

Bibliographic activity in social sciences in the era of neural networks and API interfaces

IFLA webinar
Impact of AI in Social Sciences Librarianship
3 June 2025

Digital paradigm of information activity

- Science is the most information-intensive field of activity, in which the final result of the work often depends on the quality of processing the information flow. At the beginning of the 21st century, bibliographic support for social science was carried out by many libraries, information centers and commercial services that formed numerous databases, access to which was the starting point of any research.
- Total digitalization of the processes of creating and transmitting information and the parallel development of several standards and technologies at once led to such a significant reorganization of bibliographic and, in general, information services for the scientific sphere that it is appropriate to talk about a **new paradigm** of its development.

Creation and dissemination of bibliographic information

Creation of bibliographic information

- The basis for the most profound transformations was a number of international standards adopted by academic community, allowing for the clear identification of scientific paper DOI (Digital Object Identifier) and its authors ORCID (Open Researcher and Contributor ID).
- The method of submitting scientific papers to publishers has also been unified: nowadays “manuscripts” are uploaded through standard interfaces. The author, along with the full text, enters the title of the work, its abstract, a list of references and additional sources – illustrations, tables, graphs. This process launches the entire further endless cycle of bibliographic information in the digital Universe.
- When accepting a “manuscript” submitted in this way for publication, the publisher supplements the initial information with information about the source (title, volume, issue, journal or monograph number), the pages on which the material is placed, classification indices, and assigns a DOI – a unique identifier for this work.

Dissemination of bibliographic information

- When a corresponding publication or information about it is published on the publisher's website, the DOI is automatically transferred to **Crossref** (crossref.org) – a special global agency for registering digital object identifiers created in 2000, which has turned into a giant global repository of bibliographic information continuously received from publishers.
- Crossref is an open resource, information from which can be obtained by anyone without any restrictions. To borrow data from the Network **API technology** is used.

API (Application Programming Interface)

- This software interface implements interaction between different computer programs. In particular, computer programs can remotely request and receive the necessary information from each other – in the case of interest – bibliographic information.
- Exchanges of bibliographic data, which were previously carried out exclusively manually, mainly in the process of corporate cataloging, are carried out automatically due to the API: bibliographic information is continuously transferred from resource to resource in previously unimaginable gigantic volumes ***without any human intervention***.

World wide aggregators of bibliographic information

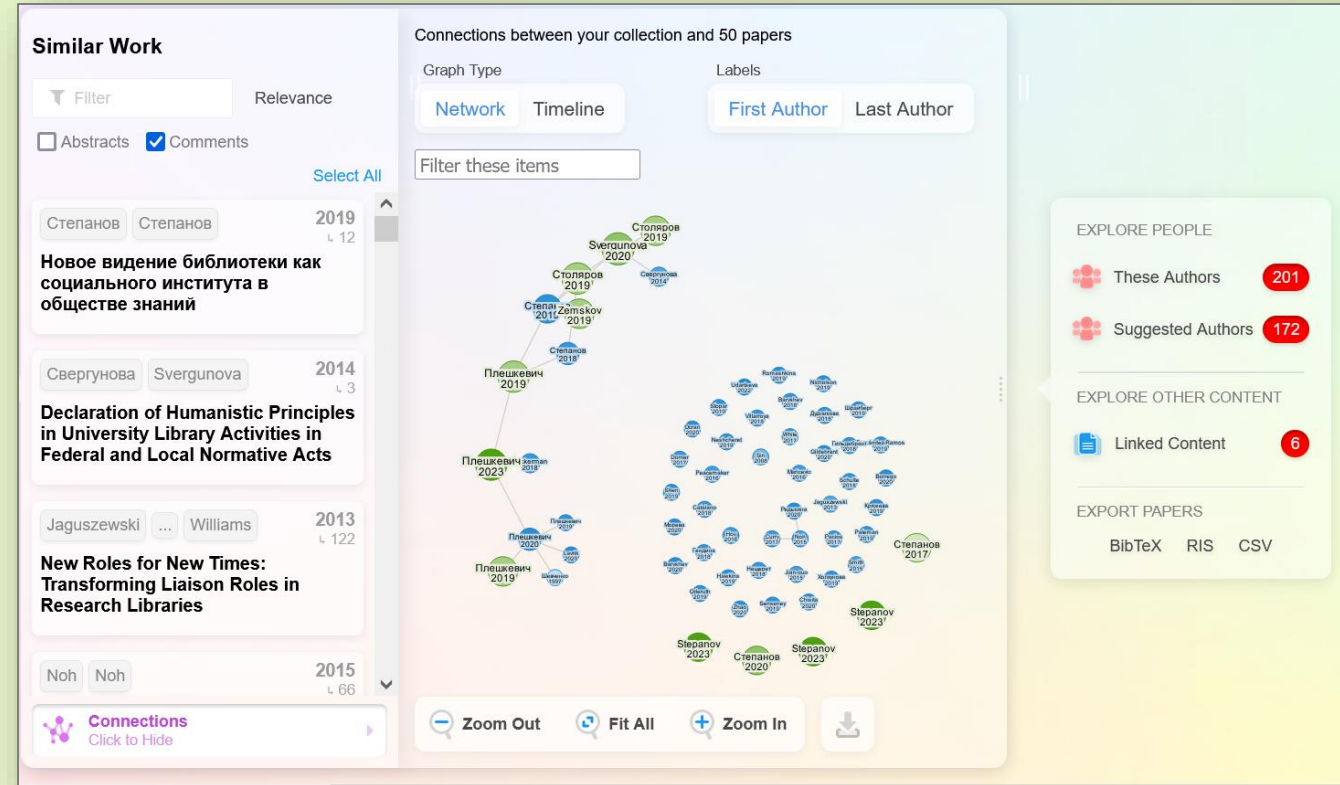
- The new types of companies, which can be described as metadata aggregators, global discovery services, or, as we see it most accurately, global bibliographic platforms, have taken advantage of the opportunities that have opened up.
- Their goal is **to unite the entire world flow of scientific and educational publications in order to provide, on this basis, a full cycle of information services to any user in strict accordance with their individual information needs.**
- In some cases, users are also offered a number of services, mostly related to scientometrics.

World wide bibliographic platforms (discovery services)

- **Semantic Scholar** (semanticscholar.org)
 - **The Lens** (lens.org)
 - **OpenAlex** (openalex.org)
 - **Scilit** (scilit.com)
 - **Google Academy** (scholar.google.com)*
 - **ResearchGate** (researchgate.net)*
 - **Scopus** (scopus.com)*
 - **Web of Science** (webofscience.com)*
- *The main task of these services is scientometrics – calculating scientific activity ratings of academic institutions and individual scientists. The Russian equivalent of such an academic service is **Elibrary** (elibrary.ru).

Academic Landscape Visualization Platforms

- The Semantic Scholar bibliographic array forms the basis for a special kind of information services that, in addition to the listed functions, include the ability to visualize the landscape of a particular scientific field.
- There are at least three similar services:
- **Connected Papers** (connectedpapers.com)
- **Litmaps** (litmaps.com)
- **ResearchRabbit** (researchrabbitapp.com)



How global bibliographic platforms work (I)

- Companies continuously, via *API technology*, query and receive bibliographic data in a standardized format (JSON or XML) from Crossref, DataCite and many publishers. Information about new publications accumulates year after year, forming an array of hundreds of millions of records to date.
- Global bibliographic platforms perform intelligent processing of bibliographic data flows, including combining information about a specific document received from different sources into a single and most informative bibliographic record, and, using neural networks, form their “semantic images” – embeddings.

How global bibliographic platforms work (II)

- Most metadata providers, including both the most famous academic brands and no less authoritative scientific repositories, willingly share bibliographic data, since access to full texts occurs only on their own platforms. The transferred metadata serve as a free advertising.
- Based on the received gigantic information array, covering, with rare exceptions, the entire world scientific publication flow, global bibliographic platforms provide **all traditional types of information support for research activities in a fully automated mode.**

Bibliographic service

Information services for scientific research

- Search for information on a one-time request.
- Implementation of current bibliographic information on a long-term request or in a non-request mode (SDI).
- Analytical processing of the document flow: compilation of bibliographic or abstract lists of a certain subject, as well as literature reviews.
- Provision of full texts of the required works.

Global bibliographic platforms successfully implement all these functions.

Global bibliographic platforms: user's view

- For the user, working with the global bibliographic platform is the most comfortable, effective and, at the same time, free way of information service. To indicate their interests, the user only needs to register and perform any actions: search on the topic of interest, upload full texts of their own scientific papers, enter into correspondence or ask questions to colleagues.
- Built-in intelligent modules, based on the analysis of the content of all user actions, form a unique profile of their interests and regularly notify about newly published works on this topic using push notifications or mailings. The user, viewing new arrivals, adjusts their request using the reactions "*Save to collection*" or "*Not relevant*".
- Users have the opportunity to form personal bibliographic collections, which are a ready-made list of literature for a new scientific publication.

New reality of bibliographic services

- Digitalization of information flows, global standardization of scientific publications and the ability to constantly exchange information between various Internet resources have led to the formation of a new reality of bibliographic services.
- Individual, even the most voluminous databases, are gradually losing demand due to the intensive development of global bibliographic platforms, which act as worldwide free discovery services for the academic community, providing both close to exhaustive completeness of the supplied data and a high level of user comfort.

The Key Problem of Global Bibliographic Platforms

- The main problem at present is the *insufficient quality of the formation of semantic images (embeddings)* of both the user's interests and the documents entering the system. Currently, search engines based on character matching often work better than systems based on meaning matching.
- An example is the low degree of relevance of the results issued by Semantic Scholar during the initial request. Only after the user selects (includes in the personal collection) relevant records, the degree of semantic correspondence of subsequent additions visibly increases.

Specialized Intelligent Services for Science and Education

- There are also intelligent services that combine information retrieval capabilities with the execution of processes characteristic of research activities, including source analysis, compiling literature reviews on the topic, extracting the main ideas from existing works and, as a result, preparing a new scientific work based on the “studied” materials. These include: **SciSpace**, **Elicite**, **Scholarcy**, **SciSummary**, **QuillBot**, **Cybertida** and dozens of others like them.
- Such services remain outside the scope of this study, since they are a kind of “subcontractors” of the systems already considered, building services based on a combination of the capabilities of global bibliographic platforms and large language models. Their impact on science and education, however, is so significant that it requires special consideration.

Conclusion

- It is obvious that information support for science will increasingly shift towards expanding the influence of global bibliographic platforms that provide users with the highest quality service built exclusively through the capabilities of software, the intelligence of which is rapidly increasing.
- Libraries around the world, especially their bibliographic services, need to take this fact into account and shift the emphasis to bibliographic processing of materials for retrospective printed collections and industry and departmental periodicals that do not fall under the category of scientific publications.

A fundamental task for library science is to determine the directions and forms of work of libraries in an era when the growing power of artificial intelligence increasingly *devalues the unique professional experience of librarians.*

The existence or disappearance of libraries as a social institution in the very near future directly depends on the answer to the question: **“Where is the place for the human brain in this artificial intellectual Information Universe?”**

Thank you for your attention!

Vadim Stepanov
inion.ru
vadimstepanov.ru