Copyright Clearance Center

RightsDirect
A Copyright Clearance Center Subsidiary

WHITEPAPER

The Top Trends in Knowledge and Information Management – and How to Get Ahead of the Curve

Phill Jones, PhD
Double L Digital

Since 1700, the global economy has gone through three radical transformations. We're now in the middle of a Fourth Industrial Revolution (4IR)[1]. The internet has led to an explosion of technology, data, and information availability that will fundamentally change the way we live, work, and perhaps even think.

Various commentators have written about this transformation in different ways. Terms like the knowledge economy[2], or quaternary sector[3], are often used to describe the rapidly growing industries that are driving these changes. In his 2019 blog post, The concept and importance of knowledge supply chains[4], Copyright Clearance Center's Chief Technology Officer Babis Marmanis puts it differently, suggesting that the industrial age is coming to a close and is giving way to something new: the information age. Semantics aside, we can all agree that a sea change has occurred. As Marmanis writes:

> ***Knowledge is increasingly the main source of economic growth, more important than land, labor, capital, or other physical resources.***

In this white paper, we examine a series of interviews with nine individuals, all key stakeholders in 4IR, most of whom work in the pharmaceutical industry. From knowledge managers to information scientists to senior research scientists, each of them provides a unique perspective on the future of data, knowledge and information management.

The most important theme to emerge from these conversations is a **recognition that integration of information across diverse sources is needed to address real-world research challenges.** For example, if the leader of a drug discovery program wants to know everything about a drug, from its pharmacodynamics to which competitors are working on it, they might ask a colleague to bring together proprietary data generated from experiments inside the company, academic research articles, and competitive intelligence information from press releases and SEC filings. In most companies, that involves searching multiple databases, sometimes owned by different groups within the organization, followed by painstaking manual synthesis. This causes time pressure and the risk of missing connections between sources.

JOB TITLES ARE EVOLVING AS BUSINESS NEEDS CHANGE

Many pharmaceutical companies have organizational silos that limit the integration of information. These silos tend to correspond to a traditional view of information and knowledge, defined by where it comes from and what form it is in. For example, implementation of electronic lab notebooks is overseen by a different department than subscriptions to research journals, even though both functions deal with the documentation of research findings.

While discussing these silos during interviews, some variation in terminology, job titles, and functions emerged, which reflects the speed at which information and knowledge disciplines are changing. Generally, *knowledge managers* are responsible for safeguarding and retaining information and know-how within a company. With a strong social aspect to their roles, they encourage tacit knowledge sharing and also have a role in preserving internal resources. In contrast, there is also a class of professionals, sometimes but not always called *information managers*, that have a more external focus; curating, selecting and assisting researchers with access to digital resources. Responsibility for resources like research data can lie with either one of the two, across both, or with neither.

There is also an increasing recognition of the importance of research data. As a result, *information scientists*, *data scientists and bioinformaticians* have become more common within companies. Sometimes these professionals are responsible for making data findable, accessible, interoperable and reusable (FAIR) (see *The Future of Fair*). These professionals can be attached to either the knowledge or information team and their functions can overlap with IT or with R&D.

This complexity of overlapping roles and functions poses a challenge when discussing modern knowledge and information management. Over time, convergence in these roles will be needed, but for now, we'll use the term *knowledge and information managers* to refer to a broad spectrum of professionals that work to support staff with all the data, information, and knowledge they need to be most effective. During one interview, a consultant in the knowledge management sector put it this way:

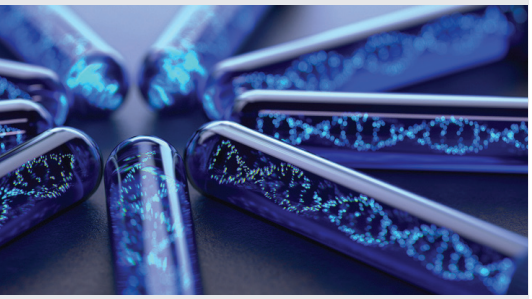
> ***"…what some people call KM [knowledge management], others call information management. To be honest, this is all just terminology. What matters is 'Do people have access to the information they need when they need it?'"***

EMERGING CHALLENGES FOR THE PHARMACEUTICAL AND LIFE SCIENCE INDUSTRIES

According to the Pharma Intelligence R&D annual review[5], the total number of drugs in the global R&D pipeline from pre-clinical development to regulatory approval hit more than sixteen thousand in 2019, an all-time high and an increase of nearly three-fold since 2001. Despite that increase, the same report tells us the number of drugs awaiting final approval or launch has actually declined over the last couple of years.

While regulatory changes are having a positive effect on time to market[6], attrition is rising. One study by Tufts Center for the Study of Drug Development found that the success rate for drugs between clinical trial and market launch halved between 2003 and 2013, down to 12%[7]. With such high attrition, it's no surprise that the cost of drug development is rising rapidly. The same study put the aggregate global cost of R&D at $2.6 billion, up from $802 million in 2003.

Taken together, these numbers point to a lot more research, a lot more information, and a lot more data to manage and organize for the same or fewer new products in the market.



According to the STM Association, about 2.4 million articles are published every year, doubling every 18 years[8].

THE EXPLOSION OF INFORMATION PUTS STRAIN ON LEGACY WORKFLOWS

In a 4IR world, there are many sources of data and information. Looking only at the most traditional measure of societal research activity — the academic journal article — gives us a sense of increasing scale. According to the STM Association, about 2.4 million articles are published every year, doubling every 18 years[8]. Increased research spending and incentives for productivity in academia are key drivers of this trend. With 60% of federal research funding in the US directed towards the life sciences, it's no wonder many researchers in the pharmaceutical industry are finding it hard to keep up to date and find relevant information.

Knowledge and information managers have been steadily increasing the number and types of sources they subscribe to or license over the last two or three decades. The professionals interviewed for this paper spoke about databases for chemical structures, clinical trials information, and drug reference sources. They also mentioned drug pipeline information, patents, drug approval packages, press releases, and even financial regulatory filings. Taken together, we have a continuum of sources telling us about every stage of the drug creation cycle, from basic science to competitive intelligence and market performance.

The number of information sources of interest will increase at an ever-faster rate, thanks to 4IR. We're seeing more non-traditional sources of data such as wearables that monitor digital biomarkers[9], open data repositories[10], population data including genetics[11], and a host of other data from novel sources.

As these new, non-traditional sources proliferate and diversify, they will create novel and difficult to predict opportunities and challenges. There's a particular challenge around integrating this data. An interviewee that works at the executive level in an information group at a pharmaceutical company said:

> ***"…most pharma companies have bioinformaticians and information scientists… I think you'll find that most companies are doing it [using non-traditional sources], and looking at more analytical ways of doing things, but it may be separate from the information group."***

As each new information type comes online, knowledge and information managers need to take a central role in defining new workflows to ingest and curate these data sources. Most importantly, they need to think about how these new data formats can be integrated without adding to workflow burdens for themselves or the researchers they support.

BREAKING DOWN SILOS TO PREVENT INFORMATION OVERLOAD

It's tempting to look at this abundance of data and information and decry the inherent information overload.

Information scientists know information overload is a misnomer. The problem is a mismatch between the volume of information and the ability to process, curate, and filter it. Multiple databases represent information silos that must be searched independently. The problem is compounded by organizational silos that make it hard to find relevant information when needed.

Silos create a risk that researchers will miss vital information. During a conversation with a senior bench researcher, they shared their experiences requesting reports from the competitive intelligence group at a major pharmaceutical company.

> ***"Sometimes I'll receive a competitive intelligence report and it will not have a key piece of information that I know for a fact is true. I'll send them an email to let them know and they apologize and add it. [The competitive intelligence group] have to compile information based on requests from any area and they don't have time to deeply familiarize themselves with the topic."***

Overcoming this challenge and ensuring an organization is getting maximum benefit from its information investments requires an integrated approach to managing knowledge and information.

Information overload occurs when the amount of input to a system exceeds its processing capacity. Decision makers have fairly limited cognitive processing capacity. Consequently, when information overload occurs, it is likely that a reduction in decision quality will occur.

— Alvin Toffler
Future Shock (1970)

A good place to start is aggregated search, sometimes called metasearch. These solutions allow information and knowledge managers to search across different databases using a single search interface. For more advanced implementations with truly integrated search, connections between different content sources can be computationally identified so vital connections are not missed.

How does aggregated search work?

The goal of aggregated search is to provide integrated search across multiple, heterogeneous sources. There are broadly two technology approaches that have been used to achieve this; federated search and web-scale indexing, but those aren't the only options.

Federated search

This approach passes the search query to multiple databases behind the scenes. Early versions used screen-scraping, which failed when a content provider changed their website. More recently web technologies like APIs have made this approach more robust.

ADVANTAGES	DISADVANTAGES
• Compatible with a broader variety of information sources. Many proprietary content providers won't allow content to be indexed.	• Can be fragile under some circumstances. • Search speed is limited by content provider systems.

Web-scale indexing

The best-known web-scale indexing service is Google. This technique involves creating a database of all the content needed to be searched. The index is searched like the index of a book and linked back to the content.

ADVANTAGES	DISADVANTAGES
• Fast and robust search creates a compelling user experience. • Computational techniques like indexed knowledge graphs can automatically surface connection between different sources.	• Not all sources can be indexed. Google, for example, can't index content behind a paywall. • Content needs to be regularly re-crawled to keep the index up to date.

A third way — fully aggregated search

Neither federated search nor web-scale indexing provides the perfect solution for a commercial knowledge and information management environment. A fully integrated approach can index content when possible and a knowledge graph of objects, concepts and connections can be created. Although the creation of real-time knowledge graphs is not computationally feasible, results that need to be retrieved in real time can be readily mapped onto a pre-existing graph. This hybrid approach can provide the best of both worlds.

CHANGES IN THE KNOWLEDGE AND INFORMATION LANDSCAPE

According to PricewaterhouseCoopers (PwC), three out of the top ten R&D spenders are pharmaceutical companies. Pharma spends an average of 17% of its revenue on R&D, with larger pharmaceutical companies tending to spend more[12]. This expenditure is for good reason: research, information, and data fuel innovation. At the same time, there is a need to make sure resources are being used efficiently.

Across all industries, workers spend more than 25% of their time looking for information[13]. That's more than one day of each employee's working week. As discussed above, big sources of inefficiency are information silos which need to be broken down. To understand how, we need to look holistically at the types of information generated or acquired by organizations.

All information is valuable whether it comes from inside or outside of the organization

Some interviewees spoke about approaches their companies had taken to aggregating knowledge and information. A common theme was an almost unrealized distinction between the internal and the external. A leader of a drug discovery center for a major pharmaceutical company described her experiences like this:

> ***"…say you're repositioning a drug and want to know everything that's been documented, you just put the molecule number in [to a search tool that we had]. There was a drop-down menu of all the things you could search. You could select all LNBs [lab notebooks], regulatory documents, all the clinical trials, but that was all internal data. The informatics people also set up another tool that was called [..] and that brought in a lot of external data sources, so if you were looking at a target you could pull in things like microarray data…"***

According to several of the people spoken to while researching this paper, over the last decade, there's been a trend towards making the most of internally generated data. Digital lab notebooks are commonplace today and many companies have invested in data warehouses and internal repositories. As one leader in knowledge management that represents an industry association put it:

> ***"One of the things that's changed is that over the last few years, there's been a focus on internal information. That's been put into the data lake, which is part of the R&D information platform. Now the focus is starting to shift towards external data. How can we really look strategically at investments we make in external data?"***

With the rise of open research and open data[14], external resources have become much more visible and available as a source of information. It's therefore important to take a holistic approach that integrates both internal and external information sources.

Only by maximizing the value that can be obtained from both, and through the synthesis of the two, can we make sure researchers can find everything they need and not miss those valuable connections to accelerate R&D.

The future of FAIR: Machine learning will take us beyond data integration

Data integration is increasingly important. The increase in momentum behind it mirrors a broader movement in science generally to make data findable, accessible, interoperable and reusable (FAIR)[15].

F indable
A ccessible
I nteroperable
R eusable

FAIR data owes its origins to open science trends in academia[16]. The concept is gaining traction and has been expanded upon by commercial companies to include internally FAIR data as a guiding principle to ensure that all results are adequately documented. As a researcher that I spoke to said:

> ***"…in our electronic lab notebook, the guidelines of how we're supposed to keep records is that anybody with a similar skill set and equally qualified should be able to follow the steps and obtain the same results."***

The philosophy behind FAIR has deeper implications than good data stewardship and reproducibility. The concept of interoperability can extend beyond data and include any class of information object. As people are becoming increasingly skilled in making data consistent, it naturally becomes easier for machines to read it. As machines become increasingly better at reading natural language and unstructured data sources, automated processing of information becomes increasingly powerful. In the long term, these two approaches will converge and there will be no meaningful distinction between the way we handle data and documents. It will all just be information.

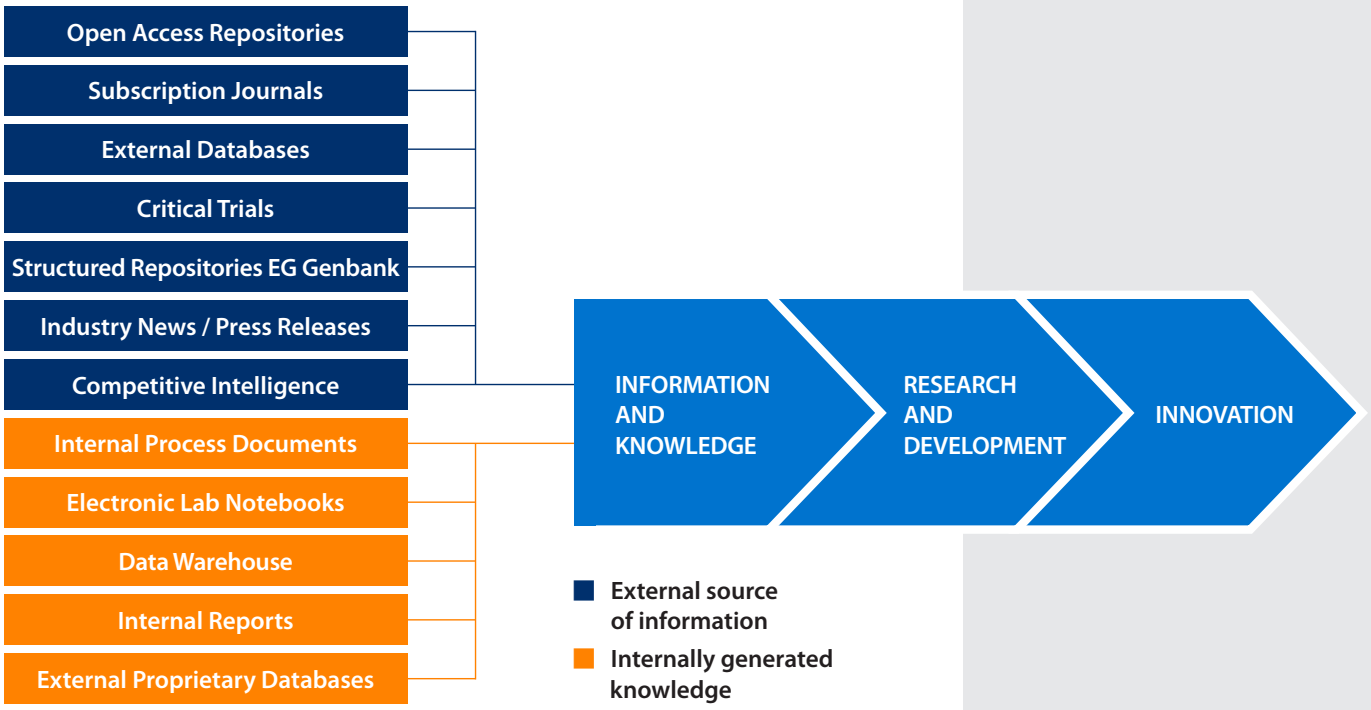
For the immediate future, knowledge and information managers should keep a close eye on semantic search approaches such as topic extraction and assertion mapping. While those approaches continue to mature, the use of curated taxonomies, full text, and fully aggregated search that cuts across databases of all kinds, literature, and document stores can significantly decrease the amount time spent searching for the right piece of information.

AI driven personalization accelerates discovery of the right information

Emerging AI technologies will play an increasingly important role in helping discover the most relevant information to any given question. At the cutting edge, some companies are using AI directly to design drug candidates[16].

For knowledge managers looking for more immediate solutions, personalized, contextualized search is both well developed and constantly improving as machine learning technologies advance. Using a truly aggregated search, connections between data and other content can be automatically identified. When cross-referenced to the search and usage patterns of the user, the algorithm can suggest or prioritize sources more relevant to the researcher.

Machine learning-based personalization has already made a huge impact in the fields of consumer retail and personalized healthcare by increasing relevance, so its widespread use in knowledge management will be a growing trend.



Innovation is fueled by a diverse combination of internally generated knowledge and externally acquired data and information.

FOUR ACTIONS THAT KNOWLEDGE AND INFORMATION MANAGERS SHOULD TAKE NOW

In the immediate term, knowledge and information managers should consider ways to use available technologies to facilitate and simplify discovery workflows.

1. The origin of a piece of information is far less important than being able to find it and make it available for the person that needs it.
2. Try to break down old assumptions and frameworks that arbitrarily demarcate information and knowledge based on format or container.
3. Consider computational techniques like machine learning that make sources interoperable even if they look very different at first glance.
4. Look for aggregation technologies that search across databases and unite disparate sources of information. Particularly as sources proliferate and more and more non-traditional sources come online, it will be necessary to use more and more automation and to aggregate information.

THE ONE THING TO KNOW FOR YOUR LONG-TERM STRATEGY

There is a longer term need to rethink how the pharmaceutical industry deals with knowledge and information. To cope with the increasing need to synthesise across many sources, it's vital to break down organizational silos that hinder the transfer and interoperability of knowledge and information. Research intensive organizations need to restructure to bring these functions together in order to rise to the challenges of 4IR and fully embrace its opportunities.

Phill Jones, PhD

Double L Digital

Phill Jones has a decade of experience bringing innovative products to market. Prior to founding DLD, Phill was the CTO at Emerald Publishing. He has had a series of roles at Digital Science (DS), including a senior role in the DS Consultancy. He also led thought leadership efforts in scholarly publishing, and developed patron driven acquisition and article syndication business models. Phill was the first Editorial Director at Journal of Visualized Experiments and is an influential thought leader in the scholarly communications technology sector. Areas of expertise include product and technology strategy, market-led digital innovation, and the changing landscape of academia. Phill is a former cross-disciplinary researcher. He received a PhD in physics from Imperial College, London and held a faculty position in neuroscience at Harvard Medical School.

[1] K. Schwab and N. Davis, *Shaping the future of the fourth industrial revolution: a guide to building a better world*, First American edition. New York: Currency, 2018.

[2] *Advancing knowledge and the knowledge economy.* Cambridge, Mass. : MIT Press, 2006.

[3] T. Selstad, 'The rise of the quaternary sector. The regional dimension of knowledge-based services in Norway, 1970–1985', *Nor. Geogr. Tidsskr. - Nor. J. Geogr.*, vol. 44, no. 1, pp. 21–37, Jan. 1990, doi: 10.1080/00291959008552242.

[4] B. Marmanis, 'The concept and importance of Knowledge Supply Chains | LinkedIn', LinkedIn, Aug. 14, 2019. https://www.linkedin.com/pulse/concept-importance-knowledge-supply-chains-babis-marmanis/ (accessed Mar. 30, 2020).

[5] 'Pharma R&D Annual Review 2019 | Pharma Intelligence'. https://pharmaintelligence.informa.com/resources/product-content/pharma-rnd-annual-review-2019 (accessed Mar. 04, 2020).

[6] A. Ward, 'Fast-track drug approval schemes signal more US-EU collaboration', May 15, 2016. https://www.ft.com/content/ef8587d4-fb23-11e5-8f41-df5bda8beb40 (accessed Mar. 08, 2020).

[7] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, 'Innovation in the pharmaceutical industry: New estimates of R&D costs', J. Health Econ., vol. 47, pp. 20–33, May 2016, doi: 10.1016/j.jhealeco.2016.01.012.

[8] R. Johnson, A. Atkinson, and M. Mabe, 'The STM Report: An overview of scientific and scholarly publishing (2018)'. STM Association, 2018, Accessed: Oct. 25, 2019. [Online]. Available: https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf.

[9] 'The Biomarker Future is Digital', Clinical OMICs - Molecular Diagnostics in Precision Medicine, Feb. 13, 2020. https://www.clinicalomics.com/magazine-editions/volume-7-issue-no-1-january-february-2020/the-biomarker-future-is-digital/ (accessed Mar. 05, 2020).

[10] M. Kindling et al., 'The Landscape of Research Data Repositories in 2015: A re3data Analysis', *-Lib Mag.*, vol. 23, no. 3/4, Mar. 2017, doi: 10.1045/march2017-kindling.

[11] 'GSK and 23andMe sign agreement to leverage genetic insights for the development of novel medicines | GSK', Sep. 15, 2018. https://web.archive.org/web/20180915123843/https://www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-to-leverage-genetic-insights-for-the-development-of-novel-medicines/ (accessed Mar. 05, 2020).

[12] PricewaterhouseCoopers, 'The 2018 Global Innovation 1000 study', PwC. https://www.strategyand.pwc.com/gx/en/insights/innovation1000.html (accessed Mar. 06, 2020).

[13] The Economist Intelligence Unit, 'The experience of work: The role of technology in productivity and engagement', Briefing Paper, 2019. [Online]. Available: https://theexperienceofwork.economist.com/pdf/Citrix_The_Experience_of_Work_BriefingPaper.pdf.

[14] D. Science et al., 'The State of Open Data Report 2019', Digital Science, report, Oct. 2019. doi: 10.6084/m9.figshare.9980783.v2.

[15] European Commission Expert Group on FAIR Data, 'Turning FAIR into reality', European Comission, Brussels, B-1049, 2018. Accessed: Apr. 01, 2020. [Online]. Available: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality.pdf.

[16] https://www.go-fair.org/fair-principles/

© 2020 Copyright Clearance Center, Inc. All rights reserved. CRP0520

Copyright Clearance Center

Copyright Clearance Center (CCC) builds unique solutions that connect content and rights in contextually relevant ways through software and professional services. CCC helps people navigate vast amounts of data to discover actionable insights, enabling them to innovate and make informed decisions. CCC, with its subsidiaries RightsDirect and Ixxus, collaborates with customers to advance how data and information is integrated, accessed, and shared while setting the standard for effective copyright solutions that accelerate knowledge and power innovation. CCC is headquartered in Danvers, Mass., with offices across North America, Europe and Asia. To learn more about CCC, visit www.copyright.com.

LEARN MORE

Maximize the ROI of your digital information investments. Contact CCC today:

info@copyright.com

www.copyright.com/rightfind-navigate

+1.978.750.8400 (option 3)