“An intensity around information”: the changing face of chemical information literacy

David Bawden
Department of Library and Information Science, City University London

Lyn Robinson
Department of Library and Information Science, City University London

Abstract
The changing nature of chemical information literacy over 50 years is examined by a comparison of a number of guides to chemical literature and information. It is concluded that: an understanding of the world of information is the sole aspect to have remained important and essentially unchanged over time; that knowledge of sources, ability to access information, and ability to organize information have been of importance throughout, but have changed their nature dramatically; and that evaluation of information has gained in importance since the advent of the World Wide Web. The link between chemical structure and corresponding substance information is the most significant threshold concept. Information literacy in chemistry is strongly subject-specific.

Keywords
Chemical information. Chemical documentation. Information literacy.

1. Introduction
Chemistry is generally understood to be one of the most information-intensive of sciences, and indeed of all subjects: "for many years, one of the most information-rich academic disciplines" [1, p.477]; “literature-heavy [with an] intensity around information” [2, pp. 19 and 25]. An ability to make good use of the information available is essential for all students and practitioners of the subject. It is conventional today to call this ability “chemical information literacy”, but the need for it was recognised long before this term “information literacy” was first used, or the concept was widely understood. It was recognised long before digital computers were applied to the storage, retrieval and communication of chemical information, and is therefore not solely associated with an ability to use the wide variety of chemical information systems now available. Indeed, Zass [3, p. 946] reminds us that problems in practitioner use of chemical information are long-standing: "sometimes it is forgotten that these 'old times' were not as good as they now seem to be, due to the fact that many chemists lacked appropriate instruction in doing these manual searches, as they now often lack knowledge and experience to make the best use of present databases."

The purpose of this paper is to examine the development of the idea of chemical information literacy over time, from the early times before the concept was described in those terms. One intention is to identify which aspects of chemical information literacy have remained constant over time, and which have emerged, disappeared, or changed with changing technologies and changing information environments. A second intention is to identify those aspects which are specific to the domain of chemistry, and those which are generic information literacy concepts which happen to appear in chemistry as in other domains. This latter is a contribution to the long-standing and on-going debate about how far there can be a generic information literacy, and how much it must always be context and subject dependent [4]. A third intention is to identify any threshold concepts of information literacy which may relate specifically to the chemistry domain. While this issue is of particular importance to academic librarians, who have been in the forefront of developing the theory and practice of information literacy, it is relevant to all information specialists dealing with subject-specific materials.

Corresponding author:
David Bawden, Department of Library and Information Science, City University London, Northampton Square, London, EC1V0HB, United Kingdom d.bawden@city.ac.uk
The first section of this paper will briefly consider the concept of information literacy itself, drawing attention to the points which will be examined later, in the specific context of chemistry. There follows an account of what may be regarded as the precursor to chemical information literacy (hereafter CIL) instruction; guides to the use of the chemical literature in pre-computer days. Then there are three ‘snapshots’ of the idea of CIL at significant points in time: around 1970, when the computer was first beginning to be used for chemical information; the early 1990s, when the computer was well-established, and the Internet was emerging; and the present day. This will involve a selective examination of CIL, and its predecessors, through the lens of a few significant sources: textbooks, articles and (latterly) websites. Use of this snapshot approach, rather than a chronological survey, is not meant to imply that nothing of significance occurred in the intervening years; it is merely a way of focusing attention on changes in CIL over time. No attempt has been made to be comprehensive in the choice of sources; rather a small number have been chosen as representative.

It should particularly be noted that this does not amount to a systematic chronological survey of the development of chemical information systems, and of their intended and actual use. That task has been ably accomplished by Willett [1] and by Gillet, Holliday and Willett [5], among others. Nor does it seek to identify what was at any stage the cutting edge of developments, particularly in computerised structure handling; that is well-documented in a number sources over the period [6-10]. The intention of this paper is to identify what was considered the norm for the laboratory-based chemical practitioner, rather than for the chemical information specialist. For example, the development, use and eventual obsolescence, of tools such as fragment codes and linear notations, covered fully in the sources above, were not a part of CIL as it affected most practicing chemists, as evidenced by the analysis below.

It is also worth noting that, for much of the period under discussion though less so at its end, the chemical information environment was very different in those industries centred on chemical research and development, and hence needing comprehensive and sophisticated access to chemical information. It was in these industries, and particularly in the pharmaceutical industry, that many developments in chemical information were pioneered; most notably structure and sub-structure searching systems, the systematic access and analysis of patent information, and the integration of internal and external chemical information sources through a single interface [11]. The nature of engagement with chemical information was therefore, at any time, somewhat different to that in other environments, although the general issues were much the same.

2. Information literacy

The term ‘information literacy’ has been used from the 1970s, to denote capabilities which were previously subsumed under ideas of ‘computer literacy’, ‘bibliographic instruction’ or ‘library skills’. The term has been developed since then to mean a set of understandings and skills to enable effective use of information, and has generated its own substantial literature; see the reviews by Bawden [12-13], Chevillotte [14], Addison and Meyers [15], Mackey and Jacobsen [16] and Stordy [17].

Typically this has been expressed in a process model, capturing a series of abilities possessed by an information literate person; a number of such models have been proposed, including two which have proved particularly influential [12, 14, 18]. The first of these is the model proposed by the American Library Association in 1989, which has spawned many similar formulations, typically as a six-stage process:

- recognise a need for information
- identify what information is needed
- find the information
- evaluate the information
- organise the information
- use the information effectively

The second is the ‘seven pillars’ model, developed by the UK Society of College, National, and University Libraries in 1999, which distinguished seven aspects, themselves relying on a basic library literacy and computer literacy capability:

- distinguish ways of addressing gap
- construct strategies for locating
- locate and access
- compare and evaluate
• organise, apply and communicate
• synthesise and create

In view of the significance of these process models, this article will consider which stages of the process, i.e. which of these abilities, have been most important in CIL at various times.

More recently, the Association of College and Research Libraries in the USA has reframed information literacy in terms of ‘threshold concepts’, fundamental ideas that define a subject, in this case information literacy, and may be barriers to understanding [19]. The question of whether any thresholds of this kind have been evident in CIL is an interesting one, addressed below.

There has been a continuing debate on whether there can be such a thing as generic information literacy, or whether each discipline or domain must have its own variant, with information literacy taking on a different ‘personality’ in different domains, with different concepts and relationships; see, for example, the analyses by Lloyd [20], by Thompson and Lathey [21], and by Pinto, Pulgarin and Escalona [22]. This article contributes to this debate by considering to what extent the components of CIL over time have been discipline-specific, i.e. of specific relevance to chemistry, and to what extent they are instantiations within the chemistry domain of general issues.

We begin by considering the forerunner of CIL: the effective use of the chemical literature before digital resources and computer searching.

3. Chemical literature instruction

The proliferation of chemical literature led to the need for guides to its use at an earlier stage than most other disciplines, the earliest substantial examples being Mason’s *Introduction to the literature of chemistry* [23] of 1925, and Crane and Patterson’s *Guide to the literature of chemistry* [24] of 1927. Of the guides published in this period, we will take for detailed consideration Malcolm Dyson’s *Short guide to chemical literature* [25], first published in 1951 and updated with a second edition in 1958. This was aimed at students and research workers, and hence more a precursor to modern CIL than contemporary texts aimed primarily at information specialists, such as that produced by the American Chemical Society (ACS) on *Searching the Chemical Literature* of 1961 [26]. It is interesting that one of the rationales for the book, given in the back-jacket publisher’s information of the 1958 edition is that “the heavy pressure on the teaching syllabuses in chemistry is such that time can seldom be spared for formal instruction in literature searching”, a thought which must find a resonance with those charged with CIL instruction today.

Dyson’s book focuses on resources, with four chapters dealing respectively with dictionaries and encyclopaedias, journals and other periodicals, textbooks and reference works, and medical chemistry reference resources. The approach in all cases is to list important and commonly encountered works, and to offer a limited amount of (sometimes idiosyncratic) evaluation: “...like most French compendia, the work lacks a good index, but is well documented” (p. 56); “...has been an invaluable desk-book ... and is recommended for general reference where a detailed search is not required” (p. 62).

A fifth chapter gives advice on literature searching; this again focuses on resources and their indexes, and the best order in which to use them, with some detailed, if dated, advice on practicalities: “keep a special set of notebooks for literature searches, and make an index record of each search. Scraps of paper are proverbially fugitive...” (p. 90). He recommends that that regular scanning searches should be performed to prepare the mind for discoveries in the literature, and emphasizes the importance of engagement with information, and urges that searching should be done by the enquirer and not delegated to others. Although he lists a few major libraries, and urges that students “cannot do better than to know thoroughly both the library and the librarian”, he rather charmingly adds that it is “only courteous and considerate to make every effort ourselves to find the required information before burdening the librarian with our difficulties” (p. 91).

With respect to domain-specific-knowledge, Dyson notes that “in searching, one necessary prerequisite is a certain agility in nomenclature” (p. 79), and that “chemical commonsense is of the utmost value in making a thorough search for subject-matter, and there are certain little tricks that only a chemically trained person can know” (p. 86): illustrating both by example of tricky organic nomenclature. The importance of substances as the fundamental concept within chemical information is implicit through the book, and is set out explicitly at one point: “It is far more difficult to search for a topic than for an individual compound. The latter has either been prepared or not; it is an entity sharply defined by its constitution and can be searched for in terms of its structure, which is unique. A topic, on the other hand, is less sharply defined and the only in terms of other concepts which may themselves be ill-defined” (p. 88). It is not hard to see this as an early statement of one of the threshold concepts of CIL.
Smith [27] gives similar examples of the necessity of chemical knowledge in terminology of substances and properties for effective searching, and of evaluation of sources: “knowing that coverage is never perfect [a searcher] learns to estimate its thoroughness for each source” (p. 21). Other chapter authors in this ACS compendium make points with some similarity to the stages of the process models of information literacy: thus Brown and Schoengold [28] state “... the importance of systematic planning of the search ... it is absolutely essential that the searcher understand the nature of the problem ... what the elements of the [search] are ... how broadly must the search be made to cover all relevant material” (pp 243 and 244), and Stephens [29] adds that “after the literature search is completed, accumulated information must be organized, correlated, analyzed and summarised if it is to be useful” (p. 263). Nonetheless, the predominant issues at this period were a detailed knowledge of a wide range of resources.

4. CIL in the early days of the computer

To represent this period, we take the second edition of The use of chemical literature [30], edited by Bottle in 1971. Its audience is the same as for Dyson’s book; students and practitioners of chemistry, though the book was, in fact, widely used by library/information students and specialists. Its structure is, again similar to Dyson, organized by forms of resources, though the categorization is much more detailed, with 17 chapters dealing with topics such as: libraries and their use; primary sources (i.e. journals and reports); abstracts; translations; tables of physical data; reference works; patent literature; etc. Some chapters take a different structure, focusing on all resources within a sub-discipline such as inorganic, polymer and nuclear chemistry. The editor noted that “the many changes in the nature and pattern of the chemical literature since 1962 have necessitated an almost complete rewriting of most of the chapters [since the first edition]” (p. v), but the chapter structure remained largely constant.

As with Dyson, the emphasis is on the understanding of content and access methods for a wide variety of printed sources. Computers had made little impact at this stage, being seen as important mainly for data. Microforms and fax were seen as significant developments, while the chapter on chemical libraries [31] suggested that “it seems clear that books, in the codex form, will continue to be the main vehicle for the storage of textual material” (p. 19). There is inclusion of what came to be known as ‘library literacy’: use of the catalogue, classified shelf arrangements, loan schemes, etc. With the benefits of hindsight, we might see this as short-sighted, but such was the general view of the time, despite initial moves to convert abstracting and indexing services, most particularly Chemical Abstracts, to digital formats.

Again as with Dyson, the focus is on knowledge of sources, particularly their coverage and indexing, and success in literature use is equated with an ability to choose appropriate sources and use their indexes. Some emphasis is placed on articulation of information need, though that phraseology was not in use at the time [32]: “in any survey or search the first requirement is a clear definition of the subject, nature and scope ... it is essential to understand clearly the purpose of the survey” (p. 254).

Little is said explicitly about the need for domain knowledge; it is assumed that the user of the book is familiar with chemical terminology and concepts, and no attempt is made to explain these when they (frequently) occur. Chemical nomenclature is mentioned at several points as an issue, and potential problem, particular when nomenclature varies over time, between countries, and between specialties. Bottle’s authors do not deal directly with this, no doubt aware that this aspect of chemical information was at this time covered by specific guides, such the fourth edition of Introduction to chemical nomenclature [33], which lamented the “great complexities and illogicalities of current nomenclature [whose] rules are as diverse and specialized as the compounds they describe” (p. 1). A familiarity of this aspect of the language of chemistry was taken for granted in texts such as Bottle’s as a sine qua non for accessing and understanding chemical information.

5. CIL in the early days of the Internet

Representing this period is the fourth edition of Bottle’s book, published in 1993 and renamed Information sources in chemistry [34]. [An anonymous reviewer has suggested that this book was widely believed to be out-dated at the time of its publication; that may be so, particularly from the perspective of chemical information specialists, but it is still gives a clear indication of what was regarded as ‘mainstream’ CIL at the time.] Digital information, or ‘non-print sources’ as they are categorized in this book, now take a major place, and it is acknowledged that personal collections are more likely to use PC software rather than index cards, but there is no mention of the Internet, which was in use for chemical purposes at the time, although the World Wide Web had not been developed sufficiently to be routinely applied to chemistry. The main arrangement of the book’s material is still by source type, but these are structured differently: into
three main sections – general sources, pure chemistry and industrial chemistry – and more by sub-discipline or purpose, e.g. inorganic and nuclear chemistry, pharmaceutical industry, than by source type. There is an emphasis on knowledge of a selection of important sources for each context, rather than the enumeration of all sources which was a feature of earlier works of this kind. Information quality is mentioned, though without the use of that term, in the sense of knowing the best journals in a field of chemistry; quality is a matter of knowing the right sources, rather than judging individual items.

Two chapters deal with online searching and with chemical structure handling, showing a new focus on use of systems and interfaces rather than sources per se, and perhaps best seen as a modern form of the ‘library literacy’, or generic information literacy, material still present in this volume. To an extent these chapters deal with generic material – retrieval interfaces, Boolean logic and the like – but rapidly become subject-specific, with structure and substructure searching, reactions databases, structure-related property databanks and so on. The subject-specific aspects seem always to be at the forefront with chemistry.

There is more of an emphasis on aids for dealing with subject specific aspects, nomenclature sources being given particular mention, and a specific acknowledgement that there are two issues, particularly in interdisciplinary aspects of chemistry [35]: “one is knowledge of the information resources available and the other is the need for background information and terminology” (p. 314). Given that both are necessarily subject-specific, this seems a clear recognition of the domain-specific nature of CIL at this stage.

6. CIL today

The present day situation is represented by Currano and Roth’s Chemical information for chemists [36] of 2014, which manages, like all its predecessors to avoid the phrase ‘information literacy’. This shows the greatest distinction from its predecessor of all the examples considered, due to the influence of the Internet, and specifically of the World Wide Web, on chemical information over the last two decades. It is still strongly focused on resources, with three main sections dealing respectively with an overview of chemical literature, with the primary literature, and with secondary sources and specialized search techniques. This is the first of the examples to use the primary / secondary / tertiary source framework as an explanatory structure, particularly valuable when a largely digital information environment removes the clues inherent in printed tools, and where the same resource may be accessible through a variety of different-looking systems and interfaces [37]. In many respects, CIL in general is now seen to have ‘caught up’ with the situation heretofore limited to the settings which pioneered modern structure-handling and integrated systems, such as the pharmaceutical industry. Systems such as SciFinder and Reaxys, which take a central place in the chapters of Currano and Roth's book, are aimed much more at the general chemist user than were earlier systems of their kind. This requires a widespread competence in their use, without the need for a specialist gatekeeper or intermediary.

While some of the chapter headings are similar to those in earlier examples, e.g. primary literature, patents, and physical properties and spectra, others are focused on systems and search techniques, e.g. searching by structure and substructure, searching using text, and reaction searching. Some of the this material deals with generic searching skills – planning the search process, use of subject databases, controlled vocabularies, citation searching, etc. - but most is specifically chemical in nature, particularly techniques for structure searching; a facility with these has to a large extent replaced the previous need for a mastery of the intricacies of nomenclature. This mix of subject-specific sources, subject-specific searching systems, and generic information skills is typical of the content of information literacy training for chemistry students reported as offered in the US in 2012 [38]; similar courses are likely to be offered in university chemistry courses worldwide, in Europe following the early example of Zass's chemical information instruction at ETH Zurich [3, 39].

The point is made that “searching for information about substances requires different access points than performing a text-based search for information on a topic of interest” [40, p.110]; an updated equivalent of Dyson’s points about substance and topic searching. Interestingly, even in this serve-yourself digital environment, with its plethora of web-based sources, the contributors echo Dyson’s advice of fifty years before to seek the help of the librarian.

While there are still lists of sources, some, such as Chemical Abstracts, Beilstein, Gmelin and Houben Weyl, familiar from Dyson, albeit now in digital form [41], there is an acceptance that it is not possible to do more choose a few representatives from the very many sources, systems and interfaces available, typically focusing on the best-known, those judged of higher quality, and those most widely, and ideally, freely, available. The contributors find it necessary to point out the value of older literature, and ways of finding it in the collections of research libraries; a point not needing to be made in the earlier example guides, where the value of the nineteenth century chemical literature was simply assumed, and instruction given on how to access it.
There is also much more emphasis than in the previous examples on the evaluation of information, a lynch-pin of information literacy, rather than simply an awareness of good quality sources. This is partly because of the very large number of web-based sources of very diverse nature, and partly because the convenience of ‘general purpose’ sources such as Google, which have greatly altered the chemical information environment. Indeed, one of the chapters is subtitled “why web search engines aren’t enough”; a sentiment which had no meaning in the earlier world of a limited, albeit large, number of subject-specific printed resources.

A very similar approach, albeit in a much shorter and summarised way with an even greater focus on a few major sources and systems, is seen in a wiki source outlining recommended information competencies for chemistry students [42]. This has four main sections: the library and scientific literature; chemical literature; properties, spectra, crystallographic and safety information; and scientific communication and ethical conduct. The first and last sections subsume the generic skills, the other two sections the subject-specific sources and systems.

7. Conclusions

Fong [38, p.14] comments, in the context of chemical information literacy training, that “there are certainly some information literacy skills that will remain timeless”. The comparison described above suggests that for chemistry the main timeless skill is to understand which resources are best to use for a particular purpose, in a subject which has always had a particularly large and diverse range of specialist resources. In terms of the process models, CIL has been, and to an extent remains, mainly about the ‘find, locate, access’ stages.

At the start of the period studied here, roughly fifty years ago the issue was to master a large number of specialised printed resources. Over time, this number has increased, most dramatically with the advent of the World Wide Web, accompanied by a need for choice of search system and interface, and influenced by competition from the well-known and convenient general search engines. We may say that an understanding of a wide range of sources remains the constant heart of chemical information literacy, certainly more so than in most subjects and arguably more so than in any other, but that what has to be understood has changed dramatically, sparking a need for a greater use of conceptual frameworks. There has also been a move towards an awareness of a few appropriate sources rather than appreciation of the totality of those available, a possibility at the start of the period, but now an unrealisable dream.

Also constant has been the need for dealing with the practicalities of information access. This again has morphed in nature: initially a matter of knowing how to use a physical library of printed materials, it is now much more about a facility with the mechanics of system interfaces. Organising information has similarly retained its importance, while changing its methods from paper notebooks and card indexes to personal digital files.

Arguably the one aspect which has remained important and essentially unchanged has been something not always included in the process models: an understanding of the world of information, as it relates to chemistry. This is close to the idea of a general ‘information fluency’, largely unaffected by technical changes, as set out by Bawden [43].

A distinct change of status can be seen in just one aspect, the evaluation of information, which has gained greatly in importance, again due to the influence of the web. Where for most of the period studied it was simply a matter of consulting reputable and reliable sources, the very richness of the new information environment makes greater demands in this respect. We might generalize this comment by suggesting that it is the web environment, and in particular the plethora of diverse sources and the appeal of general search engines, which requires CIL to ‘absorb’ general information literacy concepts and along with the subject-specific issues in a way not needed in the past; see Bruehl, Pan and Ferrer-Vincent [44] for an example of CIL training on this basis. This is certainly not to be taken as meaning that the subject-specific issues and knowledge are of any less significance than they can be shown to have been throughout the period.

And indeed with respect to the second question posed, it is clear that throughout the period an understanding of the ‘logic and language’ of chemistry has been essential. The key point which differentiates CIL from other subjects is the extent to which information is associated with, and accessed through, the concept of chemical structure. This may be regarded as the main threshold concept for CIL. Substance-related information predominates, and it is essential that this can be readily accessed, whether through nomenclature at the start of our period or through structure searching systems at its end.

As emphasised by the quotations at the beginning of this paper, chemistry has been recognised, from its inception as a modern science, to be information-intensive, in a way in which other sciences were not. With the increased use of information, and especially of digital tools and resources, in all aspects of scholarly and professional work, all scientific disciplines are now more reliant on information and data handling than previously. Some aspects of biology, in particular, are now entirely reliant upon tools for information access and processing [45]. Chemistry is arguably no longer unique in its reliance on information. However, the longevity and pervasiveness of chemistry's information
intensity is unparalleled in any other subject. To this extent CIL, though it contains some generic elements, is more strongly domain specific than any other subject. This has been both the challenge and the opportunity for CIL in the past, and will continue to be so in the future.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References


