

Search Guide

2011

This search guide serves as an invaluable one-stop resource to help information professionals migrate to the new ProQuest Dialog service. It provides key information about search, discover and analysis features, along with insight concerning query language comparisons, syntax conversions and examples of searches in various industries.

**ProQuest
Dialog™**

Contents

Background – Mission and Strategy of the Service..... 4

Search Experience Designed with Researchers in Mind..... 4

 End-users 4

 Domain-specific researchers 5

 Professional searchers 5

Supporting Search Strategies..... 5

 Search Tiers..... 5

 Breadth of Search (Default)..... 6

 Subject Industries 6

 Depth: The Database 8

 Progressive Discovery 8

 Basic Search..... 9

 Search Results..... 10

Smart Search..... 11

 More like this..... 12

 Precision Search..... 12

 Advanced Search 12

 Command Line Searching 13

 Citation Searching..... 13

 Find Similar Content..... 13

 Results Tools..... 13

 Search as part of a researcher workflow 14

 My Research..... 14

DataStar / ProQuest Dialog Query Language Comparison..... 14

 DataStar Query Language to ProQuest Dialog Query Language Translation 15

 Other differences 17

DataStar Search Syntax Conversion Guide 17

DataStar: Converting searches 17

 Operator precedence..... 17

 Boolean operators..... 17

 Truncation and wildcard characters 18

Proximity connectors 18

More proximity connector examples 19

Searching the Thesaurus 20

De-duplication in DataStar and Dialog 22

 DataStar Approach 22

 Dialog Approach 22

De-Duplication in ProQuest Dialog 22

 Alerts 22

Search De-Duplication in ProQuest Dialog 23

 Step 1: Pairing Documents 23

 Step 2: Building Clusters 24

 Preferred Duplication Option: 24

Search Solutions 25

 DISCOVER – Pharmaceutical and Biomedical Research 25

 DISCOVER – Engineering and Technology Research 26

Background – Mission and Strategy of the Service

Search and discovery on the new service is built around ProQuest's mission of being central to research around the world. As part of that mission, our aim in the development of ProQuest Dialog has been to assist information discovery for novice and professional researchers ranging from users learning the fundamentals of research to top information professionals who are charting new ground in data retrieval and management. Our goal is to provide different tools for all stages of the investigative path.

Driven by a deep understanding of the needs of knowledge workers and recognizing the increasingly interdisciplinary nature of intelligence gathering at all levels, the ProQuest Dialog service was designed to make information that historically sat across many services discoverable from one single source.

Search Experience Designed with Researchers in Mind

The ProQuest Dialog service has been constructed to support the needs of different users, moving from simple to advanced levels. Each group has unique requirements that must be addressed. This search guide serves as a one-stop resource to help information professionals migrate to the new ProQuest Dialog service, providing key information about search, discover and analysis features along with insight concerning query language comparisons, syntax conversions and search examples in various industries.

The ability to find and use relevant information to support information needs is a critical life skill. Our service helps users develop lifelong learning and information literacy, growing with searchers as they develop expertise. It solves information-gathering problems for users from multiple perspectives, from novices to information experts who want precision, as well as all levels in between. Our goal has been to design an environment in which all types of researchers can succeed in their search, discovery and evolving workflow needs.

ProQuest designed the service with three unique – but very distinct – professional groups in mind:

End-users

This user group comprises professionals with general Web experience who understand the topic(s) they want to search. They are comfortable with the Web interface and are experienced with search engines, but don't necessarily have any background with research databases. They begin learning how to do research, analyze information, and take action on it for a business need through the Internet or with guidance from their corporate knowledge/information managers, learning about timeliness, authority and bias. These types of researchers are the most frequent users of ProQuest Dialog's Basic Search form. The service was designed to be easy for them to use and serves as a solid grounding for learning basic search skills as they develop their research expertise.

Domain-specific researchers

This group typically includes professionals such as scientists, lawyers, engineers and medical specialists who have advanced degrees and can navigate through Web interfaces with their vast knowledge of the intricacies of their fields. They are experts in their major disciplines, and they are gaining proficiency, understanding and experience with the tools and information in their fields of study. They are relatively sophisticated Web users with fundamental research skills, and they have a greater sense of the role authority and bias play in selecting results. The research they undertake is driven by current projects or by the need to work with data to discover, prevent, preserve and present to interested stakeholders.

Professional searchers

These are often professional researchers, knowledge managers, or information specialists who may have advanced degrees in library science or an equivalent area. Unlike the end-user and domain-specific researcher, they are focused on understanding the resources that are available and they can provide expert searching for the enterprise staff. They often are responsible for making resources available for their colleagues to access and can help facilitate or assist in finding data that can be presented to support the core competency of the business.

ProQuest's familiarity with different types of researchers is complemented by an understanding of information-professional needs (including back-office administration systems), and a recognition of the changing role of the library as it moves from that of a repository to one of information management, technology and more. The ProQuest Dialog service provides solutions that solve researchers' problems, while also delivering administrative tools that professionals need.

Supporting Search Strategies

The New ProQuest Dialog Service

Search Tiers

The ProQuest Dialog service has been designed to offer both the breadth of searching across a range of cross-disciplinary sources and depth of information from specific sources. General search engines, or even library-discovery services, simply can't offer this combination of simple and detailed searching. For example, researchers can choose to search industries such as Pharmaceuticals & Biomedical, Energy & Environment, Chemistry and more. Within each industry, they can either search the entire industry or choose an individual database. To search more than one database and not the entire industry, they use the '**Change**' link on the top navigation bar shown on the next page to select databases by name or via industry grouping.



The search tiers start with Basic Search and offer more sophisticated search options in Advanced Search and Command Line Search. This hierarchy allows the user to run simple to complex queries, as required. Additional database-specific options are available, but in general a typical end-user does not use them. The guiding principle has been that the deeper researchers get into their work, the more they realize the capabilities for precision. Alternatively, as a search becomes broader, searchers tend to navigate to the common data elements.

Breadth of Search (Default)

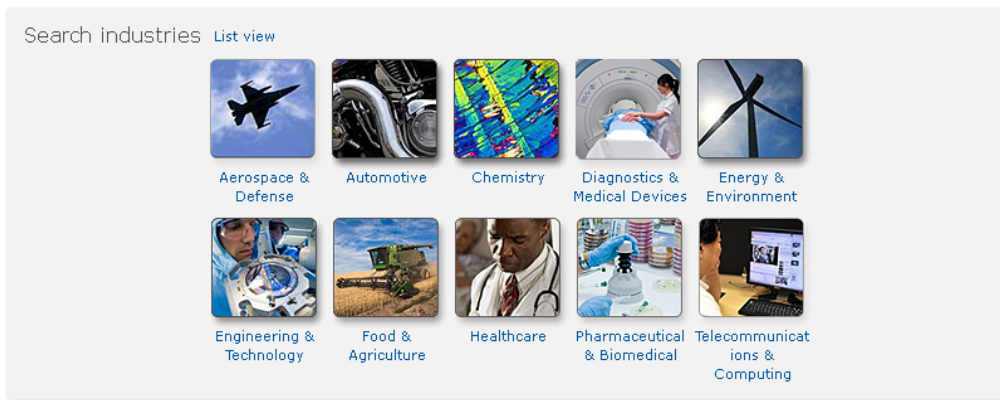
Basic Search offers the greatest breadth of search possible from ProQuest Dialog. It was designed to meet the needs of the novice and general researcher, by providing a consumer search engine-type approach; the user simply types search terms in a box to find an answer across multiple databases. The simplicity of cross-searching multiple, disparate databases in a single query without requiring knowledge of indexing or a complex syntax provides the inexperienced and general researcher with results previously only available to considerably more advanced searchers. At the same time, Basic Search does offer the ability to construct complex search queries using multiple terms with Boolean and proximity operators. This offers information professionals or domain-specific researchers the ability to present complex queries if they choose to do so.



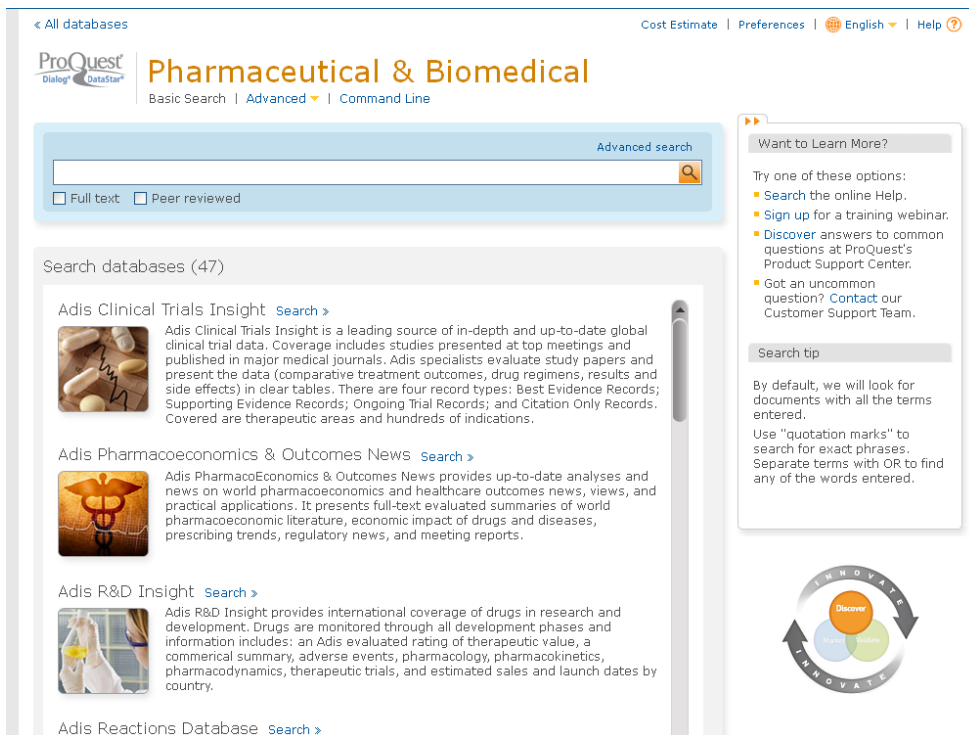
Subject Industries

Industry areas are a tool all users rely upon to search within a specific discipline, across the continuum of skill and experience. The subject industries on the ProQuest Dialog service demonstrate context and scope to users, and clearly show what is available within a given set of databases. They visually articulate the type of content likely to be found from these sources.

A key component of the service is these pre-defined subject industries. However, recognizing the uniqueness within the professional market, ProQuest Dialog also offers an immense amount of flexibility through the ability to customize subject areas. Administrators who have access to the ProQuest Administration Module (PAM) can create multiple-disciplinary subject areas that include the core databases end-users and domain-specific researchers can search.



The industry groupings offer breadth of content for researchers in a given discipline while offering enhanced precision for end-users as they refine their research skills. It also helps researchers to associate a specific database with a given subject.



The subject industries also serve to broaden the “content pool” for domain-specific researchers – often used to accessing a single database – who need to find content from adjacent disciplines as their work becomes increasingly multi-disciplinary. For example, after finding articles concerning side effects of a specific drug within the Adis Clinical Trials Insight database, a researcher may want to find more evidence from other

authoritative biomedical databases such as MEDLINE®, BIOSIS Previews®, Embase® and SciSearch®: A Cited-Reference Science Database to further support their findings.

Depth: The Database

Domain-specific researchers will usually know where to find the content or the databases they are interested in searching. ProQuest Dialog allows users to drill down to see individual databases covering specific disciplines. Single database search pages still offer a Basic Search box for end-users, with visuals and other context-providing clues to the scope of the database in the form of pictures and specific information. In addition, by displaying industries and their databases, ProQuest Dialog can ensure novice searchers and professionals alike will not find themselves “lost” within the new service.

Importantly, individual databases also offer greater detail for the specialist who recognizes and understands the characteristics of that database, such as the expert indexing and editorial input associated with each database. From the Advanced Search page general or domain-specific researchers can obtain the level of detailed granularity within their area of interest. Users can search using additional limit options such as content type, publication, date-range selection, classification and more, depending on the database, to allow the general or domain-specific researchers to drill down to the particulars they need.

Progressive Discovery

The search experience is designed to accommodate a wide range of expertise, from the basic keyword searching of an end-user to the complex query syntax of an expert. Features such as search tips, Auto Complete, Smart Search and faceted results found in the “Narrow results by” options enable researchers to build queries and get the results they want. The interface is

intuitive enough for end-users to perform their searches by simply entering terms and either pressing the Enter key or clicking the search icon for retrieval.

Basic Search

It is expected that end-users of ProQuest Dialog will conduct their searches using the Basic Search page. Given that this is the access point for the vast majority of end-user searchers, Basic Search is designed to be entirely inclusive. It has been engineered to be both comprehensive and deep, offering search (recall) of all metadata elements (e.g. author, title, abstract, etc.) for all documents in the databases covered. A Basic Search, whether across all available databases, in a single subject area, or in an individual database, searches across all title, author and subject fields (including descriptors for databases previously on DataStar®), controlled vocabulary terms, locations, classifications, geographic terms, etc. This functionality ensures that the value of subject indexing in research databases is maximized, even for end-users unfamiliar with indexing terms. In addition, the growing body of user-contributed content – Tags, Shared List Titles and Shared List Descriptions – can be accessed via Basic Search.

Basic Search covers every informational element known in the document including some highly specialized fields in the full record that are not searchable fields – for example “related works,” and full text of the document (excluding scanned PDF without Optical Character Recognition or ASCII behind the image), where available. ProQuest Dialog continues to offer broad field searching in the Basic Search in order to give the novice and general researcher access to all available metadata without requiring specialist knowledge of database field codes. The service helps these researchers “cut through the clutter” of larger numbers of results by also providing an advanced relevancy ranking engine that automatically surfaces the results most pertinent to the user’s query.

The advanced researcher can enter elaborate syntax such as field codes, Boolean operators, proximity connectors and truncation without leaving Basic Search.

Basic Search syntax

Searching on ProQuest Dialog follows common search conventions, which implies the AND operator between words, and use of quotation marks to search for an exact phrase. This is a key distinction from some of the Dialog legacy products such as DialogSelect or DialogWeb, where two words were automatically searched as a phrase. This change to common search conventions in the new service has been driven by researcher expectations at all levels, including domain-specific researchers, as well as by the desire to provide a more uniform search experience.

Auto Complete

As users key in their search terms, the Auto Complete feature suggests successful search queries the researcher may wish to use. This feature solves several very common problems facing users. First, typing takes time. Second, some users do not spell well. Third, some users may need help identifying relevant search terms. In usability tests, this feature has proven valuable for end-users and some domain-specific researchers struggling to construct

appropriate queries. Again, this follows common Web conventions and no learning curve is involved.

Note: Auto Complete can be turned off by clicking an option at the bottom of the Auto Complete drop-down box. It can also be turned on or off via My Research and PAM.

Lemmatization

A researcher's search term is processed by the ProQuest Dialog search engine, which looks to find the broadest set of relevant results for the term. One way to achieve this is by using lemmatization. Lemmatization is the process of reducing a word (in this case, a search term) to its root form, taking into account its context, meaning and part of speech, and then returning results based on the search term and its inflected forms. Lemmatization differs from stemming in the way that the root form is obtained. Stemming, in its various forms, attempts to find the root form of a word without regard to context or part of speech, and therefore may not accurately retrieve many inflected word forms and words that have multiple meanings.

- ProQuest Dialog also analyzes synonyms driven by a synonym dictionary to return highly relevant results by finding variant spellings for contemporary language – color vs. colour (US vs. British).
- ProQuest Dialog uses lemmatization to enhance recall by mapping search terms to: obtain singular, plural and other forms of the terms – *nurse*, *nurses*, *nursing*, for example
- Verb forms are excluded

The use of lemmatization reduces the need for learning specific search syntax such as truncation and nested Boolean queries. For the inexperienced and general researcher, this means the search engine recalls all pertinent documents, yet remains precise in delivering relevant content. If domain-specific or general researchers know exactly what they want, they can use quotes or precise syntax to generate a concise search.

Note: Lemmatization and Synonym expansion can be turned on or off via My Research and PAM.

Search Results

Search results are provided with relevance as the default sort order, with options to sort by publication date in either descending or ascending order

Relevance

The ProQuest Dialog search engine enhances the precision of a search through its relevance-ranking engine, which determines the order in which results appear. We have worked to ensure the right balance of weights on certain elements to meet relevance expectations of the three researcher communities. The engine weighs a variety of factors when determining the order in which results appear, including:

- **Number of query term matches**—For queries with multiple terms that are not joined by AND, the more of the individual query terms that match a document, the higher the ranking of a document. Terms that include truncation characters do not contribute to relevancy-rank calculations.
- **Proximity of terms**—Documents with multiple matching terms are ranked higher if the terms are in close proximity to one another. Known phrases are treated separately and given a higher weighting than individual terms. Proximity weighting also takes into consideration the distance between terms relative to the beginning of a searchable field.
- **Frequency of terms**—A greater number of occurrences of an individual matching term raises the document's overall rank value. Terms that are very common across all documents such as *a*, *an* or *the* have no rank value and do not contribute to overall document rank.
- **Context of terms**—Different key fields within a document are configured to carry a higher relevance weight when they contain matching terms. The following fields in ranked order from high to low are: title, subjects, abstract, author and then full text. Publication date is a secondary sort applied when two or more results have identical relevancy weights. While publication date is more heavily weighted in other platforms, the sorting provided by placing greater weight on term frequency and proximity in the new service ensures that results that most closely match the terms searched appear first. If the searcher wants currency over relevancy, the option to sort by descending date is also available.

Did you mean?

'Did you mean' assists with misspelled words. If you get no results, the search engine searches for one alternative spelling and will include a statement showing you what was searched. If the original search term has some results, those are presented along with suggested alternate spellings.

A word about stopwords

The new ProQuest Dialog service analyzes the entire universe of data in the fields searched; extremely common words (*and* and *or*, for example – often used as stop words on other search engines) will not contribute to the overall document rank by their frequency. These terms are, however, factored into the search through their proximity to other search terms to give the researcher the most relevant results possible. So, for example, if you search for *meaning of life* the frequency of the term *of* will not contribute to the rank of one document against another because it is such a common term, but its proximity to *meaning* and *life* allows more relevant results.

Smart Search

The Smart Search feature taps into ProQuest Dialog's powerful indexing to supercharge results, analyzing a user's search, and then offering suggested topics at the top of the results screen in certain databases. Smart Search was designed to help researchers who aren't comfortable extracting controlled vocabulary terms from a thesaurus. Technology developed by ProQuest

analyzes a user's query, maps the terms to the controlled vocabulary and then offers suggestions for related topics.

More like this

Along the same lines as Smart Search, the ability to "See similar documents" based on the content of a given item is built into every ProQuest Dialog document available on the new service. On every document view page, users will be able to find similar or related items through a new feature called "More like this" that appears as one of the tools on these pages. Similar to Smart Search, this feature leverages the structure and content of the document being viewed to find and recommend more items along similar content lines, saving searchers time and effort once they have found documents valuable to their searches.

Precision Search

Advanced Search

The Advanced Search feature is targeted for the general and domain-specific researchers who want the power of searching with precise syntax and field codes without requiring actual knowledge of those specific tools. To provide an easy transition from Basic to Advanced search, the user can obtain identical results by entering queries in the first row of the Advanced Search form as from entering the same queries in the Basic Search form. As users become more familiar with the functionality, they can quickly progress to more advanced features found in the intuitive Advanced Search interface. ProQuest Dialog provides structured content with metadata fields searchable individually, and the service offers a range of advanced query tools and options. Because there are multiple fields with syntax choices in each row of advanced search, it is possible to create a different search depending how items are keyed and conjoined.

In each row in Advanced Search, the researcher can select from a list of fields to qualify to those that are specific to the database(s) selected. Many of the ProQuest Dialog databases have specialized search fields available in drop-down lists when searching individual or grouped databases. Those specialized search fields that are not available in a drop-down box are still searchable with field codes and can be located by consulting the ProQuest Dialog ProSheets for each database (<http://www.dialog.com/prosheets>). Novice and general researchers can be overwhelmed by too many options, so domain-specific researchers' needs for full access to all options are accommodated through field codes. The default selection in Advanced Search is to search all fields (except full text) as in Basic Search. This arrangement provides a greater selection of results for the user while the enhanced relevancy sorting enables the user to zero in on key results.

As researchers progress and require increasingly advanced syntax, ProQuest Dialog provides support for this ongoing development of information literacy through, for example, help text, on-screen instruction and context-sensitive search tips. Readily-accessible browse lists support general researchers.

As previously mentioned, multiple search terms are queried with an implied AND between terms, such that searching *washington state budget* is the same as searching *washington AND state AND budget*.

Command Line Searching

Command Line searching, used primarily by professional searchers, is also available. For those professionals who prefer to build more inclusive search strategies while having the ability to combine sets as needed, the Command Line search form enables continuous searching.

Look Up Citation

The ProQuest Dialog Look Up Citation search is designed for finding known documents. It includes the most common metadata elements such as title, author, publication title and year.

Find Similar Content

Another feature cited in many user surveys as highly desirable is the ability to construct a query by example, i.e. to provide a highly useful or relevant content item to a search interface and have that item be the basis for finding other items with similar content. While such a feature was not possible to implement with older search technologies, the search engine of the new ProQuest Dialog service is able to analyze a large text document via natural language processing algorithms and conduct just such a query. The new service provides this feature via the “Find Similar” search form, and it allows users to copy and paste any amount of text into the search form, submit it for analysis by the ProQuest Dialog search engine, and receive suggested search results based on that analysis. While this is a relatively new technology in the library marketplace, it is generating interest because it allows searchers to use a relevant document, passage, discussion or similar topic as the basis for finding the content that drives further discovery.

When you click **Search**, ProQuest Dialog evaluates the text, identifies what it determines are the key terms, and returns a search results list containing similar documents.

Results Tools

Researchers at all levels want to get to relevant content quickly. Novice and general researchers typically don’t want to refine and re-execute searches. Following widely-accepted Web conventions, they prefer to “drill down” from a wider pool of results. In response to this preference, ProQuest Dialog provides a set of tools to help researchers with the search they have executed and ensure the results integrate easily into their typical workflow.

Usability studies show that the simpler the search query, the more likely the researcher will benefit from the tools provided on the Search Results page. These include suggested subjects generated by **Smart Search**, **Narrow results by** filters and **Search within results**, simple tools to modify searches as well as sorting options to view results from different perspectives.

Search as part of a researcher workflow

My Research

My Research offers a means for researchers to collect and organize their research into folders, manage Alerts and RSS feeds for monitoring purposes, re-execute Saved Searches to find new content, use tags to organize documents by topical areas and create Shared Lists for dissemination and publishing. It is designed to support the workflow of all researchers, but it is particularly valuable for domain-specific researchers to keep up-to-date with the latest research and communicate their work to a wider audience.

DataStar / ProQuest Dialog Query Language Comparison

The capabilities of the DataStar Query Language (DQL) and the ProQuest Dialog Query Language (PDQL) are very similar. The most frequently used DataStar query operations have exact equivalents in PDQL. The following table shows the proportion of queries that use each of the DQL operators for Alerts and interactive sessions. The “PDQL Equivalent” column indicates how close a match the corresponding PDQL operator provides. Two of the operators which PDQL does not support are used very rarely in DataStar (SAME and XOR). (Note that the percentages listed may add up to more than 100 because many queries use more than one operator.)

DataStar Query Operator Usage

DQL Operator	Alerts %	Interactive Search %	PDQL Equivalent
OR	42.4	8.2	Exact
ADJ	28.0	28.9	Exact
AND	26.6	48.0	Exact
'-' (bound term)	22.0	8.5	Close
WITH	5.5	0.6	
NOT	3.7	0.6	Exact
NEAR	3.5	0.4	Exact
NEXT	0.9	0.1	Exact
SAME	0.4	0.1	
XOR	0.0	0.0	

The most noticeable differences to a DataStar searcher using ProQuest Dialog will be a different and larger set of field names, different query syntax and different strategies for expressing some very specific information needs. DataStar documents are limited to 32 fields that are labeled with two letter codes. Because of the limited number of fields, some fields contain several pieces of information. For example, the Source field (SO) may contain the journal name, publication date, page references, ISSN, etc. In contrast, ProQuest Dialog

documents can have any number of fields and the field codes are usually two to six letters long, although they can be any length. Document mapping from a DataStar field name to the corresponding ProQuest Dialog field name is therefore not always exactly the same.

DataStar also supports a search feature known as Quick Codes which are MeSH and Emtree qualifier abbreviations and database-specific index and limit options that can save you keystrokes while searching. In ProQuest Dialog MeSH and Emtree indexing provides Quick Code groupings that allow users to search multiple subheadings with a single abbreviation, such as QX for Quick toxicology.

DataStar Query Language to ProQuest Dialog Query Language Translation

This table summarizes the similarities and differences between the DataStar and ProQuest Dialog query languages.

DataStar Operator	ProQuest Dialog Equivalent	Notes
AND, OR, NOT	Same	
ADJ	Surround with double quotes e.g., heart ADJ attack becomes "heart attack" Use PRE/0 heart pre/0 attack	Terms must be together and in the order specified. Note: When using quotes lemmatization is turned off but with pre/0 lemmatization is still active.
NEXT	PRE/5	Terms must be within five words and in the order specified. Note: In ProQuest Dialog the bare word 'pre' is treated as a query operator.
NEAR	NEAR/5	Terms must be within five words and in any order. In ProQuest Dialog the word <i>near</i> entered by itself is treated as a search term, e.g., <i>heart near attack</i> will search for three terms with an implied AND between the terms: <i>heart AND near AND attack</i> . Apply the syntax NEAR/# between terms to ensure the search engine reads your strategy correctly.

DataStar Operator	ProQuest Dialog Equivalent	Notes
'-'	No exact equivalent, but surrounding with quotes is close	<p>"Bound" term</p> <p>In DataStar, using a hyphen creates a single bound-phrase, usually descriptor, search term. This allows you to search for a term such as <i>coronary-artery-disease</i> as controlled vocabulary. This is also used in author searches (e.g., find <i>smith-j</i>, but not <i>smith-jones</i>) and with descriptors such as MeSH or Emtree terms, where multi-word keywords are joined with a hyphen.</p> <p>ProQuest Dialog ignores all punctuation so the queries are equivalent and will both match documents containing coronary-artery-disease.</p>
WITH	<p>Two cases:</p> <ul style="list-style-type: none"> • Full text: NEAR/8 • "Structured field" restriction: partially supported 	<p>Terms must be in the same sentence in any order.</p> <p>When searching full text, using NEAR/8 or NEAR/10 will approximate using WITH, however the search will also match documents where the terms occur in separate sentences.</p> <p>See below for discussion of "structured field" searching.</p>
SAME	No equivalent	<p>Terms must be in the same field in any order.</p> <p>Usage is negligible in DataStar.</p>
XOR	(x AND NOT y) or (y AND NOT x)	Usage is negligible in DataStar.
Truncation: \$ and \$N	?, *, [*n] and \$n	Most existing uses of truncation are to find multiple word forms (e.g., <i>infect</i> , <i>infected</i> , <i>infection</i>). Since searching for variant word forms (lemmatization) is the default behavior in ProQuest Dialog, truncation may not be as heavily used, but is available for any cases that are not covered by lemmatization.

Other differences

There are two other differences between DataStar and ProQuest Dialog that affect searching:

- The default query operator used when no operator is specified between two terms is different in DataStarWeb, DataStarClassic and Alerts, and ProQuest Dialog. DataStarWeb uses ADJ, DataStarClassic and Alerts use OR and ProQuest Dialog uses AND. The query *oil exploration* is interpreted as follows by the three systems:

Search System	Interpretation of <i>oil exploration</i>
DataStarWeb	<i>oil ADJ exploration</i>
DataStarClassic and Alerts	<i>oil OR exploration</i>
ProQuest Dialog	<i>oil AND exploration</i>

- DataStar searches all fields by default whereas ProQuest Dialog searches only title, subject, abstract, author and text.

DataStar Search Syntax Conversion Guide

The ProQuest Dialog service brings the highly-regarded Dialog and DataStar professional search engines together, combining them into a single, streamlined search experience. As a result of this consolidation, there are some differences relating to the operators, truncation characters and field codes. This guide explains the differences.

Note: In the following examples, T stands for (search) *Term*.

DataStar Search Syntax Conversion Guide

Operator precedence

Operator precedence refers to the order in which terms joined by operators in search queries are interpreted by ProQuest Dialog.

DataStar	ProQuest Dialog
(), ADJ, NEAR, NEXT, WITH, SAME, AND, NOT, OR	(), NEAR, PRE, NOT, AND, OR

Boolean operators

DataStar	ProQuest Dialog
T1 AND T2	T1 AND T2 T1 T2 (space defaults to AND)
T1 OR T2	T1 OR T2
T1 NOT T2	T1 NOT T2

T1 XOR T2

Can be emulated with:
(T1 NOT T2) OR (T2 NOT T1)

Truncation and wildcard characters

DataStar	ProQuest Dialog
Term\$	<p>Term*</p> <p>Truncation can be used on the right side, left side, or inside of a word, retrieving from zero characters up to a maximum of 10 characters currently. It will retrieve up to 500 expansions (word variations).</p>
Term\$#; Term\$2; etc.	<p>Term[*#]; Term[*2]; Term[*5], etc.</p> <p>Term\$#; Term\$2; Term[*2], etc.</p> <p>Limited Truncation – to indicate the maximum number of characters to be included in the search – make sure to either add the square brackets and the number after the asterisk or use \$ followed by the number. This will include from 0 characters up to # more characters. The maximum number of characters is 10. Term[*10].</p>
Search example: cat or cats or catty	<p>Term OR Term? OR Term??</p> <p>One question mark '?' will retrieve one character only (but not zero), e.g., <i>cat?</i> will retrieve <i>cats, cate, cato</i>, but not <i>cat</i>. Two question marks '??' will retrieve two characters only, but not less than two e.g., <i>cat??</i> will retrieve <i>catty</i>, but not <i>cats</i>.</p>

Proximity connectors

Some DataStar proximity connectors are not directly supported in ProQuest Dialog. Terms to emulate them are suggested below.

DataStar	ProQuest Dialog	Notes
T1 ADJ T2	"T1 T2" T1 PRE/0 T2	Use of quotation marks turns off the automated plurals and alternate spellings feature. To retain automated plurals and alternate spellings, use: T1 T2 or T1 PRE/0 T2.
T1 NEXT T2	T1 PRE/5 T2 T1 P/5 T2	<p>PRE/# allows any number of words between T1 and T2, with T1 and T2 in the specified order.</p> <p>PRE/5 replicates the DataStar NEXT operator.</p> <p>Used alone, PRE defaults to PRE/4.</p>
T1 NEAR T2	T1 NEAR/5 T2 T1 N/5 T2	NEAR/# allows any number of words between T1 and T2, with T1

DataStar	ProQuest Dialog	Notes
		<p>and T2 in any order.</p> <p>NEAR/5 replicates the DataStar NEAR operator.</p> <p>Used alone, NEAR alone defaults to NEAR/4.</p>
T1 WITH T2	<p>T1 NEAR/# T2</p> <p>T1 NEAR/15 T2</p>	<p>DataStar supports the concept of the same sentence, and the WITH connector specifies that T1 and T2 must occur in the same sentence, in any order. On ProQuest Dialog NEAR/# defines the number of terms that can appear between T1 and T2. There is no direct equivalent of WITH on ProQuest Dialog but the recommended suggestion is to use something like NEAR/8 or NEAR/10, or use your best judgment of the proximity of your terms</p>
T1 WITH T2	<p>T1 LNK T2</p> <p>T1 -- T2</p>	<p>The WITH connector in DataStar also links terms in the same subfield in particular fields, such as Descriptor term plus Subheading (qualifier) in a Thesaurus search or elements in the same row in a table meta field, such as Development Phase and Indication in drug pipeline databases. ProQuest Dialog introduces the operator LNK (or --) between the same elements, e.g.: "<i>phase III</i>" WITH <i>psoriasis</i> will become "<i>phase III</i>" LNK <i>psoriasis</i>.</p> <p>See Searching the Thesaurus below for more examples.</p>
T1 SAME T2	T1 NEAR/150 T2	<p>DataStar supports the concept of the same paragraph, and the SAME connector specifies that T1 and T2 must occur in the same paragraph, in any order (a paragraph being either a text paragraph or an indexing field). In ProQuest Dialog NEAR/# defines the number of terms that can appear between T1 and T2. There is no direct equivalent of SAME on ProQuest Dialog but the recommended suggestion is to use NEAR/150.</p>

More proximity connector examples

DataStar	ProQuest Dialog
T1 ADJ T2 ADJ T3	<p>"T1 T2 T3"</p> <p>Note: Quotation marks turn off the automated plurals and alternate spelling.</p> <p>Use T1 PRE/0 T2 PRE/0 T3 to retain automated plurals and alternate spelling.</p>

DataStar	ProQuest Dialog
(T1 or T2).AB.	AB(T1 OR T2)
(T1 or T2).AB, TI.	AB(T1 OR T2) OR TI(T1 OR T2) AB, TI(T1 OR T2)
(T1-T2).DE.	DE.EXACT("T1 T2") DE.X("T1 T2")
(T1-T2).AU.	AU.EXACT("T1 T2") AU.X("T1 T2")
(T1 OR T2) WITH (T3 OR T4) (T1 OR T2) SAME (T3 OR T4)	(T1 OR T2) NEAR/15 (T3 OR T4) for the same sentence (T1 OR T2) NEAR/150 (T3 OR T4) for the same paragraph
T1 OR T2 ADJ T3 OR T4	T1 OR "T2 T3" OR T4

Searching the Thesaurus

Many databases in ProQuest Dialog are indexed with a hierarchical controlled vocabulary (Thesaurus) that can be consulted online while searching either in one or more databases. The following table shows how to search using the Thesaurus in ProQuest Dialog compared with the similar experience in DataStar.

DataStar	ProQuest Dialog	Notes
cattle-weighers.DE.	SU("cattle weigh*") MESH(T1) EMB(T1)	The general field code (Field tag) for Descriptor in ProQuest Dialog is SUB or SU. MEDLINE and Embase have special field codes for the MESH terms (MESH) or Emtree terms (EMB).
cattle.W..DE.	SU.EXACT("cattle") SU.X("cattle")	EXACT will search for the specified term or phrase only, excluding descriptors containing more unspecified terms. E.g.: SU.EXACT("cattle") will exclude SU("cattle weighers"). Shortcut of .x(TERM) has been added for .exact("TERM"), where TERM is the thesaurus term, e.g., cab.x("abdominal

DataStar	ProQuest Dialog	Notes
		surgery").
T1#.DE.	SU.EXPLODE(T1) MESH.EXPLODE(T1) EMB.EXPLODE(T1) MESH#(T1) EMB#(T1)	Exploding the term will automatically include in the search all the narrower terms under the specified descriptor in the Thesaurus hierarchy. Similar to DataStar, the explosion can be selected as an option in the Thesaurus window, or searched directly by adding the specific tag in Advanced Search, Command Line Search or Basic Search.
Abdominal-neoplasms#.MJ.	MJEMB.EXACT.EXPLODE("Abdominal Abscess") MESH#(Abdominal Abscess) MJMESH.EXACT.EXPLODE("abdominal neoplasms")	Similar to DataStar, the Major Descriptors in the Thesaurus can be searched separately either by selecting the option in the Thesaurus window, or by adding a specific field code in Basic, Advanced Search or Command Line search: MJMESH for MEDLINE and MJEMB for Embase. Command line syntax: explode. Shortcut of mesh#(TERM) has been added for mesh.exact.explode("TERM")
(abdominal-Cancer WITH DI).DE.	PHS("PHASE III" LNK PSORIASIS) EMB("abdominal cancer" -- "diagnosis") MESH("abdominal-neoplasms" LNK "diagnosis") MESH("abdominal neoplasms" LNK DI)	Similar to DataStar, linking a descriptor term to a Subheading (Qualifier) can be done by selecting the proper qualifier in the Thesaurus window, or by using the connector LNK (or --) in Basic, Advanced, or Command Line Search. Quotation marks are optional.
(Abdominal-cancer-DI#.MJ.)	MJEMB.EXACT.EXPLODE("abdominal cancer -- diagnosis") MJEMB.EXACT.EXPLODE("abdominal cancer" -- di) MJEMB.EXACT.EXPLODE("abdominal cancer" LNK "diagnosis")	Similar to DataStar, the search in the Thesaurus can combine the explosion, the search as Major Descriptor and the link to a qualifier. This can be done either by selecting the related options in the Thesaurus window or by using the proper tags and connectors in Basic, Advanced

DataStar	ProQuest Dialog	Notes
----------	-----------------	-------

Search or Command Line Search. Be sure to add quotation marks when using .EXACT. expressions.

De-duplication in DataStar and Dialog

DataStar Approach

- Compares only the first 150 characters of the titles.
- DataStar considers the title, first title (e.g. translated title), and author in scientific, technical and medical records; the title and source in business content.
- Alert deliveries are stored in clusters/folders for duplicate identification and removal.

Dialog Approach

- The titles and authors of each article are processed to remove the additions that an information provider might add to the original title and to remove differences in abbreviations, spelling and punctuation.
- Next the "processed titles and authors" are compared. If identical, the records are considered to be duplicates.
- If there is either a difference in the titles or the author names that was not removed during the processing, the difference would prevent the records from being considered duplicates.
- For example, the difference of square brackets in one title and parentheses in the second title may not be enough to prevent the titles from being duplicates, but the difference between I and L for an author's middle initial would be enough to prevent the records from being identified as duplicates.

De-Duplication in ProQuest Dialog

Alerts

- De-duplication is applied against a single Alert or an Alert using multiple databases.
- The default setting is to have duplicate items removed.

- For each Alert, an “Alert History” file is maintained, keeping track of all previous documents sent for that Alert.
 - The “Alert History” file currently has no time limit for storing the history of delivered documents
- When new documents are generated by an Alert, the Alert system checks the Alert History file to see if this document, or any duplicate of this document, has ever been sent before. If it has, the document is not sent again.
- Documents generated for an Alert are checked for duplicates using “Quick De-duplication” which checks the following field values for matches:

For Most Documents	For Patent Documents
<ul style="list-style-type: none"> • Publication date • Important words from the title <ul style="list-style-type: none"> ▪ Very common and very rare words are dropped to help normalize small differences between different versions of the document. • Important words from the publication name 	<ul style="list-style-type: none"> • Country • Year • Patent Number (normalized) • Kind Code • A patent family record is never allowed to be considered a duplicate of any other record.

Search De-Duplication in ProQuest Dialog

Step 1: Pairing Documents

- Pairs of documents are chosen for comparison based on shared values among certain combinations of the following fields:
 - Author
 - Year
 - Volume
 - Normalized publication ID
 - Start page
 - Title
 - Important words and word pairs chosen from the document's bibliographic information overall

- A set of reference words is used that have been collected from citation/reference bibliographic data
- These words are sorted by frequency, with important words chosen from the middle of the range, i.e. skipping over the terms that are too frequent to contain useful information, or too rare to be of much use

Step 2: Building Clusters

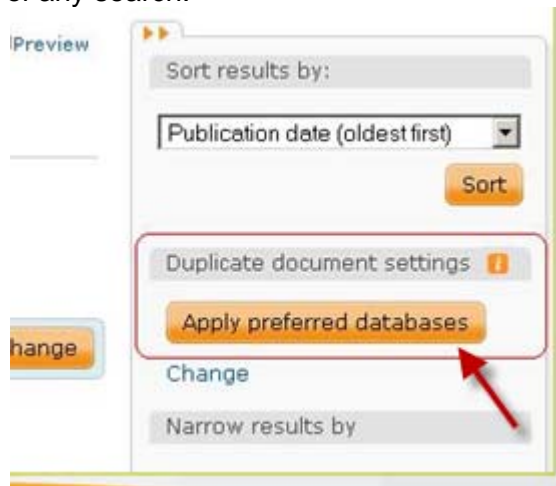
- Chosen document pairs are compared using all available bibliographic data
- Those judged to be sufficiently similar, based on measures of term overlap, are clustered together.
- For large clusters we enforce more strict constraints on overall similarity

Step 3: Applying the Algorithm

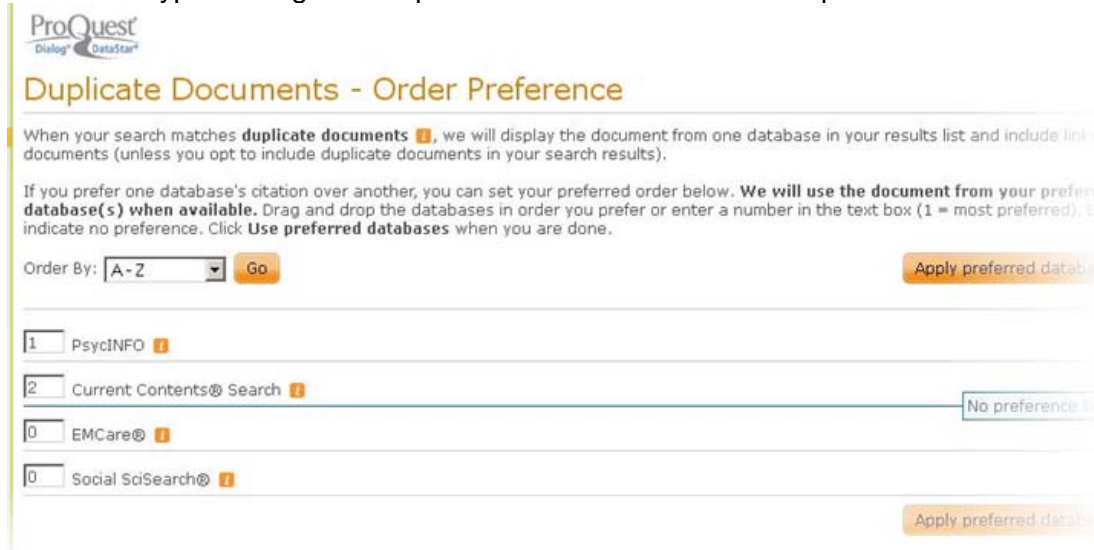
- A set of constraints is applied to each cluster, based on analysis of the following fields:
 - Year
 - Publication date
 - Title
 - Author
 - Volume
 - Issue
 - Start page

Preferred Duplication Option:

A user can rank databases in order of preference and apply those preferences to the results of any search.



- Any documents in the results set with duplicates from a preferred database are replaced.
- A user can type or drag and drop to rank databases in order of preference.



- Preferences are retained for the session, but a user must choose to apply those preferences to each set of search results.
- The results list then indicates that the substitution has been made for this search.

Search Solutions

DISCOVER – Pharmaceutical and Biomedical Research
Search for a biomedical subject
Find reported adverse effects of a particular drug
Find a specific citation using the Citation Look Up feature
Use the MeSH online thesaurus in MEDLINE®
Use drug link subheadings in Embase®
Search for drug names in Embase®
Locate clinical trials being conducted in a particular year on a specific drug
Search for new drug approvals (NDA) and abbreviated new drug applications (ANDA)

DISCOVER – Engineering and Technology Research

[Search for an engineering subject](#)

[Search for a comprehensive list of an author's publications in an engineering field](#)