QL2 manageable data | on demand

Turning Unstructured Information into Actionable Market Intelligence – *QL2 Technology & Solution Overview*

WHITE PAPER

Situational Background

Business today is moving faster than ever. The Internet has changed the shopping and purchasing habits of consumers in every sector. Despite the increasing volume of data available to guide marketing and business decisions, marketers still rely on information that is not only too broad and expensive, it is often received too late to be useful.

While this type of information is publicly available on the Internet, a large portion of it exists in unstructured form – for example, in Web pages or online documents – rather than in traditional database formats. As a result, extracting the data and translating it into actionable intelligence presents a difficult challenge.

QL2 bridges the intelligence gap by providing the technology to automatically gather and present the most current, relevant and comprehensive view of the marketplace. It empowers companies with the information to compete effectively, regardless of size or resources.

What is QL2

QL2 is a flexible, intelligent technology solution that actively gathers structured and unstructured information from virtually any networked data source – across the Web and internal company sources – and delivers it on demand in a highly intuitive, real-time interface. By collecting and categorizing specific information according to pre-defined rules, QL2 provides timely, actionable intelligence that allows companies to make better decisions faster.

QL2 is uniquely designed to collect and organize unstructured, uncontrolled data – that which exists outside a company's four walls – from websites and other complex network environments. QL2 can access information from virtually any document format, including data hidden behind forms, embedded in graphics and contained within metadata. This requires sophisticated navigation and semi-structured document manipulation capabilities.

QL2 transforms this raw data into market intelligence, providing actionable metrics and qualitative information on what is being sold anywhere, at any time. It is the definitive source for market intelligence from the consumer's perspective.

Designed for simplicity, QL2 does not require the installation of complicated hardware or software. In fact, minimal involvement from the IT department is needed to set up and run QL2. With only a few lines of code written by a moderately skilled developer, QL2 can accomplish the same tasks as hundreds of lines of code written by a Java engineer. Also, since QL2 operates within a small technical footprint, it is unobtrusive to the data sources it accesses.

Intelligent agents perform the data collection. These are small, yet powerful pieces of software code that collect information from targeted sources. Much more efficient than Web crawling, this technology is highly scalable and capable of handling one query or thousands of queries with minimal CPU load. After years of constant use and improvement, QL2 has made the customization and refinement of data extraction commands a simple task. In addition, QL2 streamlines the data acquisition process through simple mapping and lightweight normalization.

QL2 Use Case: Pharma Competitive Intelligence

A top-five global pharmaceutical company uses QL2 to collect broad types of market intelligence and to organize its internal content. Specifically:

QL2 collects and categorizes broad sets of market intelligence – encompassing public, protected (user ID and password) and private information including:

- Clinical trial status and adverse events information from FDA, CROs and clinical trial-specific websites.
- Formulary data from hospitals, medical groups and healthcare insurers.
- News directly from competitors and industry press.
- Detailed technical information from grants, research and conference proceedings.

QL2 aggregates purchased content and collected market intelligence into a centralized repository, allowing researchers to directly access this information through simple queries. A simple keyword search UI tool was built using QL2.

QL2 Compared to Other Technologies

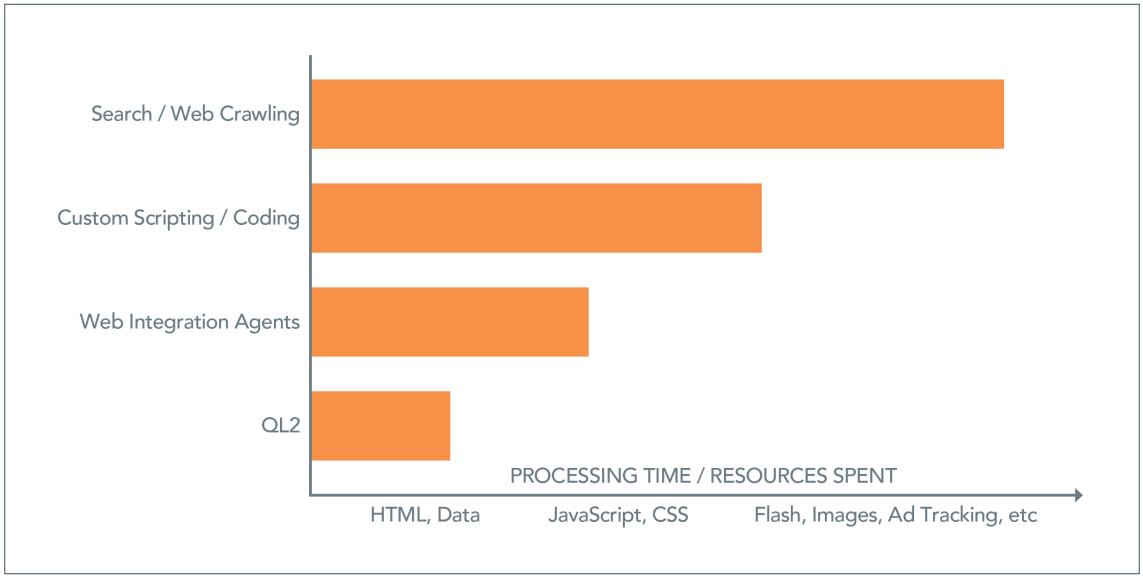

Diagram 1 – QL2's approach is dramatically faster than the alternatives since it leverages HTML and data elements that load on web pages first

When considering how to extract and organize information from the Web and other sources, technologists face many choices. Traditional messaging middleware as well as data integration and programming tools each offer a solution. However, these technologies face many limitations when compared to QL2's ability to efficiently access unstructured information from uncontrollable data sources.

Technology	Limitations for Unstructured, Uncontrollable Information	QL2 Solution Benefit
ETL / EII	• Built for data warehousing • Often not possible for real-time information • Requires domain knowledge and developer skill sets across each source technology	• Agnostic to target source of data required • Designed to access information in real-time
EAI / Messaging Middleware	• Built for messages, not data collection • CPU intensive when performing large-scale data collection • Requires expensive developer skills	• Operates within a small technical footprint • No special developer skills required
Search / Web Crawling	• Built to collect everything, so it captures a lot of useless data • Cannot access protected and private information (behind forms, passwords, etc.)	• Collects targeted information • Unobtrusive to data sources • Designed to access public, protected and private information

Table 1 **(continued on next page)** *– Comparing Technologies*

Technology	Limitations for Unstructured, Uncontrollable Information	QL2 Solution Benefit
Web Integration / Agents	• GUI-based agents inefficient for targeted data extraction • Difficult to scale	• Built for enterprise scale • Intuitive development environment without GUI limitations
Scripting / Coding	• Custom approaches result in long development cycles and one-off solutions	• Automates manual scripting, reducing time dramatically

Table 1 **(continued)** *– Comparing Technologies*

QL2 Architecture

QL2 technology is the underpinning for our on demand services and is available as packaged software for on-premise installation. Both deployment styles are built on the same underlying stack of technologies. QL2 technology is broken into four major areas:

- **QL2 Agent Platform**: Automates the collection and categorization of unstructured information.
- **QL2 Service Platform**: Transforms information into market intelligence.
- **QL2 Storage**: Stores the terabytes of data collected for QL2's on demand services.
- **QL2 On Demand Services**: Offered directly to business users to answer specific marketing or business questions.

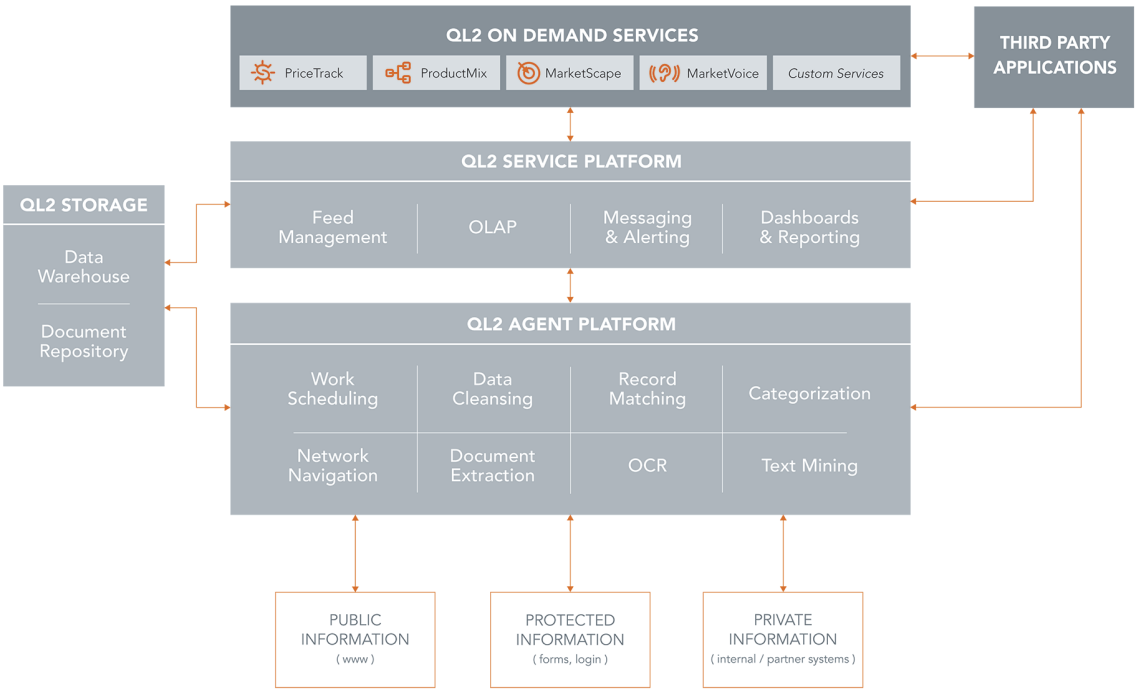

Diagram 2 – QL2 On Demand Services

As illustrated in the above diagram, the QL2 Agent Platform collects information from a variety of sources including websites – both public and password protected – as well as file systems, data sources and enterprise applications within a customer's firewall. The information gathered is organized and presented in a QL2 on

demand dashboard or exported for integration into a wide range of external applications. QL2's Service and Storage Platforms manage the activities of agents and the storage of the data they collect, respectively. This information is presented to the customer through QL2's On Demand Services, in both vertical- and horizontal-specific products.

The Agent Platform

The Agent Platform is designed to automate information extraction by collecting and organizing public, protected and private information. Agents intelligently extract information from any information source, and are developed in QL2 Studio. Elements of the Agent Platform include:

Work scheduling: This provides agents with instructions for gathering data, such as at what time, and how frequently, the information is to be collected.

Network navigation: QL2 can navigate to any data that is accessible by a person through a Web browser or other desktop application. QL2's technology can handle all standard Internet/Web protocols (including HTTP/HTTPS for Web data, ODBC for databases and many other specialized protocols such as FTP, IMAP and several email protocols). Agents can supply valid credentials to access information that requires authentication.

Document extraction: QL2 technology understands dozens of common file formats, including HTML, XML, PDF, Microsoft Word and Excel, RTF, CSV and many others. It is also able to pull data from compressed or archived formats such as GZIP and ZIP files.

Data extraction: A core feature of QL2 is its suite of dozens of procedures for pinpointing data in documents. These range from extremely low-level processes (extraction-based regular expression patterns) to high-level, knowledge-based techniques (extracting items such as company names, addresses, dates, times and geographic locations based on natural language processing).

QL2 Use Case: Content Aggregation for Educational Publisher

A leading educational publisher uses QL2 to process millions of pages of PDF content from lesson planners provided to teachers as an addendum to textbook sales.

- Information varies by state (e.g., 3rd Grade Mathematics in California), and encompasses millions of pages of content in aggregate.
- Content is extracted, normalized and categorized to perform content verification and editorial tracking; also provides valuable audit trail.
- It is too expensive to gather data from the source – using QL2 is dramatically cheaper and faster.

QL2's deep understanding of tabular structures lets agents apply their instructions to whatever data format has been accessed, including CSV or Excel spreadsheets as well as tables in HTML or Word files. A QL2 agent programmer gives the agent an instruction such as "select all values from the columns named 'Ticker Symbol' and 'Price'" (encoded as a particular programming language syntax). The exact same programming instructions can be used to extract data from many different native data formats. QL2 also has sophisticated PDF processing capabilities that can recognize table structures as well as font and style information, which simplify data extraction from PDF documents in many cases.

Data cleansing: Once data has been pinpointed, it must undergo a variety of cleansing and normalization processes, such as currency conversion, date and time reformatting and street address corrections. QL2 technology offers a variety of tools for data cleansing and normalization, from hand-coded rules embedded directly into the agent program to the ability to call out to external services (for example, using the U.S. Postal Service address normalizer to correct U.S. street addresses).

QL2 also has a sophisticated data typing and normalization mechanism that allows developers to specify constraints on data (for example: "must be an integer between 10 and 100" or "must either be null or the value LCD or Plasma"). Agents can use this mechanism to validate data at any point during execution:

- While reading input configuration information (to ensure that input parameters are valid).
- While generating intermediate data during processing (as a debugging tool to ensure agents are behaving properly).

- While generating output data when the agent is finished (to ensure that the expectations of subsequent consumers of the data will be met).
- **Object matching:** QL2 agents are often used to extract descriptions of "objects" (products, companies or people) in order to aggregate information about a single object obtained from different sources. A fundamental task in this process is to determine whether two object descriptions from disparate sources refer to the same object. The table below illustrates the need for advanced logic to determine whether two patient records refer to the same object (top), or two different ones (bottom).

Clinic	Preferred Provider	Matched
Robert Rollins 1410 SW Lake Mist Rd. Kennesaw, GA 30155 (404) 462-2482	Robert J. Rollins 1410 Southwest Lake Mist Road Kennesaw, Georgia 30155-4253 404-462-2482	
Cheryl Lynn Anderson 11837 Outer Bank Pkwy. Lexington, KY 41028 (502) 895-0116	Cheryl Lynn Anderson 183 Outer Bank Place Lexington, Kentucky 41028-5549 502-897-7828	

Table 2 – Object Matching

QL2's Object Matcher technology solves this problem in two phases. In the first phase, data values for each field are compared using field-specific rules. In the second phase, the field-specific comparisons are aggregated to calculate the overall similarity of two descriptions by filtering items based on column-specific thresholds and assigning different strengths to each column.

In comparing patient data, for example, phase one would measure the difference between two birth dates or the distance between two ZIP codes; phase two would then define the importance of the age difference relative to the location as well as all other data fields. A match or a discrepancy is declared based on the decision weight placed on each field.

Categorization: Another common operation of QL2 agents is the categorization of objects by predefined labels. For example, if an agent is examining a financial services company's website to find pages that list top company officers, then the categorization task may be to label each page with one of the following: "Leadership page", "Likely to lead to a Leadership page" or "Useless". Then, the agent will selectively crawl pages from the "Likely to lead to a Leadership page" category until the "Leadership page" is found. QL2 can perform categorization by using either hand-crafted heuristics or a machine-learning based text classification algorithm.

OCR: Using Optical Character Recognition (OCR), QL2 agents are able to extract data – such as a product name or price – that is only available as text embedded in an image. Because agents can read text represented as images in web pages and PDF documents, previously inaccessible data is now completely available for extraction and delivery.

Text mining: QL2 has several capabilities related to Text Mining and Natural Language Processing. Not only can QL2 identify objects in text, such as people and companies, it can also arrange text into predefined categories, such as news and blog article topics. In addition, QL2 can:

- Summarize long documents into a few sentences based on the relevance of the sentence to the important topics within the document.
- Cluster a set of documents into groups according to topic.

- Analyze document for sentiment (for example, identify passes that are positively or negatively disposed toward a particular product or brand).

The Service Platform

The Service Platform is where the information gathered by QL2 agents is transformed into actionable intelligence. Elements of this platform include:

Feed management: Enables users to schedule and configure feeds – sets of information gathered by agents at regular intervals. This function provides users with a high degree of customization and automation when running regular information collection tasks.

OLAP (online analytical processing): This element provides a structure for reporting the data. It pulls data from multiple sources and organizes it to provide a result.

Messaging and alerting: The messaging function provides a very concise summary of the last 10 hours of data gathering, while the alerting function communicates degrees of behavior over a fixed period of time. For example, QL2 can send an email when certain thresholds – such as prices and product availability – are met.

Dashboards and reporting: The dashboard provides an on-screen reporting of the market intelligence that agents have learned over a long time span.

QL2 Use Case: Order Application for Automobile Dealerships

An Automotive Solutions Provider developed a custom vehicle order application using QL2. QL2 delivered a dramatically simpler and cheaper application environment for the provider compared to traditional app-dev alternatives. This solution:

- Guides the user through complex manufacturer-specific vehicle configurations [make, model, color, options packages], collecting orders from hundreds of dealerships and submitting aggregated orders to manufacturers.
- Easily handles complex order configurations, including complex java applets.
- Integrates with three major automobile manufacturers' systems.
- Is used by hundreds of automobile dealerships.

The Storage Platform

In an on demand world, storage is a critical component of any data gathering technology. QL2's Storage Platform includes the following elements:

Data warehouse: Collects all extracted data in a secure, high redundancy environment for easy access.

Document repository: Serves as a virtual warehouse that stores documents accessed during the data extraction process, such as reports or website screen shots.

On Demand Services

QL2 On Demand services are hosted on QL2's On Demand platform, which accesses and analyzes over 500 million individual data points each month and supports more than 250 clients. The platform includes a variety of services, and also allows for customization based on client needs. On Demand services include:

PriceTrack: Captures pricing information from nearly any website that consumers use to shop and delivers that information to clients on demand. This service provides QL2 clients with an instant snapshot of pricing and promotional activity across the competitive landscape. In addition, it enables clients to verify that their own pricing strategy is being properly implemented across channels, and allows them to view competitor promotions, purchasing restrictions and product availability.

ProductMix: Lets clients determine whether their category management strategies are working by comparing products and categories against those offered by the competition. With ProductMix, clients can view their product performance across multiple channels, conduct internal analysis and identify market trends.

MarketScape: Provides information about important news and trends by canvassing company, industry and market news, publications, competitor websites and regulatory and compliance filings.

MarketVoice: Allows clients to monitor consumer discussions and reviews about the companies, brands and products in their industry. MarketVoice tracks, quantifies and interprets consumer sentiment by unlocking the user-generated content in various forms, formats and graphics that often hide this valuable information from other research methods.

Custom Services: The flexibility of QL2 technology allows it to be easily customized to capture and organize any type of external market information from publicly available sources. There are several useful applications for this type of customization:

- **Compliance:** Check public data sources to verify whether the company is in compliance with local, state and federal laws.
- **Procurement and logistics:** Automatically shop for lowest prices on common goods and services to lower the company's overall operational costs.
- **Channel management:** Ensure that downstream partners (distributors, resellers, retailers) are properly executing the company's programs.

On Premise Services

QL2 offers its software to clients who desire to maintain market intelligence on their own premises to maximize privacy and internal control. QL2's On Premise Solutions leverage the same platform as QL2 On Demand, and include tools to create, automate and analyze data collection and integration activity. QL2 technologies are readily deployed within a client's firewall and feature three levels of application:

Developer Efficiency

QL2 is a more efficient – and therefore more cost effective – development environment for extracting and organizing unstructured content.

- QL2 provides simple scripting and navigation tools for identifying and extracting content.
- QL2 is purpose-built to perform unstructured data extraction, automating tasks in a few lines of code that would require hundreds of lines of Java code.
- QL2 developers are less expensive to train than Java developers since it is a simpler task.

QL2 Studio: The basic desktop product intended for customers who want to do their own in-house agent development. QL2 Studio includes all the features and benefits of QL2 technology as well as a helpful debugging environment. It also executes scripts.

QL2 Server: Designed for large-scale agent execution, QL2 Server is for clients engaged in high-volume, ongoing data integration activities. QL2 Server is scalable, allowing script execution to be coordinated concurrently across several machines, limited only by available hardware and network bandwidth.

QL2 SDK: QL2 agents can be embedded in a client's own applications using QL2 SDK. For example, a client can develop an application that invokes QL2 technology and then returns the retrieved data to the application for subsequent processing. QL2 technology can be called from COM, Java or C++ application programming interfaces. Because QL2 agents have a very small memory and CPU footprint, they are easily embedded within a wide variety of applications.

Conclusion

By easily extracting and organizing data from virtually any website, network environment or document format, QL2 turns data into market intelligence, providing actionable metrics on what is being sold anywhere, at any time.

QL2 uses extensive HTML support that facilitates the collection of data points from unique online sources, using customer- and industry-specific pieces of software code known as agents. Because this technology is more efficient than Web crawling, it is scalable from one query to thousands with a minimal CPU load.

QL2 is available to clients both as an on demand service delivered in real-time through the Internet, and as on-premise software. In addition, it can be customized to meet specific company needs.