



Abrevity SLICES Data To Extend Information Value

Analyst: Anne MacFarland

Management Summary

All businesses use information for administration, logistics, operational reporting, and similar office routines. They may be shocked by the volume of email that they now must handle, or the size and complexity of their customer databases, but how they *use* information has not changed over the years. However, there is a subset of these businesses for which information *is* the business. These are found not just in the expected areas of education, media, and entertainment, but in design houses and research organizations that serve almost every industry. **For these information-intensive companies, changing how they use their critical content, and how they leverage their reuse of it, is key to maintaining profitability in the face of stiff competition.** These enterprises face some particular and acute data pains they are feeling, which generally comprise the following three.

- 1. Like all organizations, they have a case of data glut.** But, in their case, curtailing the glut is ruinous to the business model. The “all”-ness of critical content has value. More is better.
- 2. Capturing or creating the “upstream” information that constitutes most of critical content takes time and costs a lot.** In oil and gas, for instance, the information needed to locate deposits is expensive to glean and expensive to analyze. Research is a huge expense for the drug and medical industries, and the regulatory reporting they must do add still more costs. The tradition, based back in the days of paper, of gathering fresh data for each project adds time to market, something we all bear in drug and medical costs.
- 3. There is great difficulty in knowing the extent and character of the aggregates of critical content they already have.** The information is vast, and changes constantly. Comprehensive knowledge of what you have is not satisfactorily supported by search, where you find only what you are looking for. What is needed is a comprehensive discovery and characterization of all that is in a domain, together with the tools to make the discovered attributes the basis of policies and actions in a way that does not slow process or system performance to a crawl. And, this last requirement is the major challenge.

Abrevity, Inc., based in Cupertino, CA, focuses on the needs of information-centric enterprises. Its solution scales to the vastness that is required, and is priced to accommodate platoons of workers using the same content in different ways. Their *SLICEBase* avoids the limitations of relational databases. Its discovery mode of classification exceeds the grasp of search. Please read on for more details.

IN THIS ISSUE

➤ Information as a Process Asset	2
➤ Abrevity SLICE	2
➤ Abrevity Products	3
➤ Go-to-Market Particulars	4
➤ Conclusion	4

Information Value Management

The value of information is often in the aggregate. Think of the information that underlies sports statistics financial balance sheets, and medical research. Consider the data collected by medical instruments. Such aggregate information comes as a multitude of instances, or cells, each of which may have a few or hundreds of data points. In scientific inquiries, scientists use instruments to capture and interpret data. Any particular scientist will analyze only a few aspects of the data captured. It would be ideal to capture the entirety of the data, and its context, on some more shareable medium (networked disk or tape), from which it could be recalled to an instrument (the only device that can interpret it) for the next researcher wishing to analyze different data points from the same data. The challenge is to preserve all of the context that determines when the second use of data is appropriate, without creating such an operational logjam that reuse becomes ineffective.

- **Relational Databases** classify information into table structures. They focus on relations between entities (not a strength needed by “upstream” data). Their need for consistent formats, schemas, and uniformly populated categories would require standardization – but what constitutes the context of information varies widely, and precision of relevance is important in information-centric enterprises. If a traditional relational database could be constructed to contain all the potential contextual metadata elements of upstream data, operations on it would be insupportably slow. With the constant change of information, rebuilds would be crazy.
- **Object structures** can accommodate any number of attributes, but analyzing objects scales even more poorly than relational databases. Moreover, you are talking about many elements here. There must be an extraction of attributes that can be manipulated to discover, quickly, all the relevant information for a particular purpose, as these purposes arise.
- **File systems** can accommodate almost anything and, these days, can scale very large – but they don’t tell you much. Like the card catalog of a library, they tell you some basic characteristics, and where something is stored.

- **Search** is a relatively new form of ad-hoc data organization, but it, too, has limitations¹. Like a corrective lens, search cannot be a one-size-fits-all. Search spiders are sent out with limited intelligence and limited missions. Based on their returns, search creates an index, or several indices, that take up storage resources. And, they don’t tell you broadly what is out there.

What the business built around critical content needs, to extend the value of its information, is a combination of the *classification* of databases, the *completeness* of objects, and the *comprehensiveness* of file systems, and the *opportunism* of search. This finding information, this metadata, must be in an environment that is neither prescriptive nor restrictive, and one that is easy to comb through.

Abrevity SLICE

SLICE, an acronym for *Scalable Linked Independent Customizable Entity*, is Abrevity’s patented granular file attribute discovery vehicle. A SLICE entity can be an attribute, or it can have attributes, or both. It is inherently flexible in nature – a scale-free structure.² Some information will map to multiple entities, and some will be more limited. The ability to go characterize information at a granular attribute level permits appropriate information reuse, which is what gives this approach great business value.

SLICE attributes are not limited to what can be discovered by parsing and analysis. They can be endowed with metadata from taxonomies, or keyword lists by users, if that is what is needed to enhance the enduring value of the critical content.

Abrevity SLICES are buckets, not meaningful containers. This is a great advantage, because it means that they can divide, like cells, as they grow large. Processing the information of value can be quick and transparent, not a painful ingestion or transformation process.

¹ For a more thorough treatment of search, see **The Clipper Group Explorer** entitled *Enterprise Search Adds a User Dimension to Business Information Organization*, dated July 28, 2005, and available at <http://www.clipper.com/research/TCG2005048.pdf>.

² The main characteristic of scale free networks is that its nodes vary from connection-rich hubs to more sparsely connected nodes in a peer-to-peer non-hierarchy. Other scale-free structures include social networks the World Wide Web, and airline routes.

Abrevity *SLICEBase*

Abrevity SLICE management is accomplished by a non-relational database called *SLICEBase*. The *SLICEBase* engine discovers files, extracts information with which to classify them to the extent desired, manages their placement on tiered storage, and reports to administrators. Here is how it works.

SLICEBase has the ability to scan a volume, directory, instrument, or any other network share. It then parses all the information from the path, by use, and by directory. It extracts all the relevant contextual information (including a time tag) from these structures, and expresses it as metadata³. It can, if desired go further and extract more value based on the content of the file, or the method used by the instrument to collect the data – to whatever granularity is needed to maximize the ongoing utility of the information. This will vary. A traditional search engine can be embedded in the *SLICEBase* engine if such a capacity is useful. Each *SLICE* in the database has its own schema, and can read the information on other slices even if the schema is not the same.

Archiving retention schedules are accomplished through a tag. In an appropriately-attributed form, information can be moved off of instruments onto other storage, and even moved⁴ off-line onto tape. *SLICEBase* can expire by deletion or by removing the link to the data, whichever is preferred. *SLICEBase* can also time stamp the link, adding knowledge of process evolution that would not otherwise be captured.

Could this be applied to other kinds of business data, like customer data and product data, as a more flexible bucket? Yes, but product data and customer data need consistency and coherence first, something better served at this point by relational database characteristics. The *upstream* of critical content is where the pain that Abrevity can satisfy is greatest. It is, at

³ Abrevity also extracts access control information from file systems, databases, and applications and stores this as tags as well. Thus, Abrevity can enforce policies like retention and security based on the discovered file and path attributes of the information assets, rather than creating another set of policies

⁴ Abrevity prepares the information for export and checks the data after it is moved, but leaves the actual movement to third party software. A stub is left, so the Abrevity *SLICEBase* has a pervasive and detailed awareness of the whole domain of critical content.

present, a remarkably inefficiently-used and badly-managed asset, and in information-focused industries, those inefficiencies demand a solution. Abrevity offers the following products targeted at this need.

Abrevity Products

File Data Manager

Abrevity feels that, without content and context visibility into data sources, Information Lifecycle Management, or ILM, is limited in what it can accomplish. Compliance, security, and legal discovery, all basic rationales for instituting ILM, all require a completeness of purview to be effective. Content management and publish and subscribe approaches enforce prescriptive usage of a defined set of information assets. For any organization seeking to get the most use out of their information assets, a more flexible, discovery-focused approach to unstructured file data may be more satisfactory.

File Data Manager uses the *SLICE* method and the *iBase* engine to extract all the file path metadata (far more information than that given by file extensions). An advanced version analyzes file content where desired. The software resides of laptops, desktops, and/or servers. It is priced by the capacity of the data it addresses. There is no extra charge for client instances, so an enterprise can deploy instances of the software (either basic or advanced, depending on the role of the worker) to all knowledge workers using the data, greatly leveraging the cost effectiveness of the product.

File Data Manager produces metadata attributes, including those to enforce security, retention, backup, and migration. With the file content analysis capabilities available in the advanced version of the product, it can produce content summaries. In the process, *File Data Manager* identifies duplicate files, outdated files, and certain kinds of at-risk content.

Abrevity's file data manager uses no agents. It can deal with both *CIFS* and *NFS* files. Its entity approach incurs less storage overhead than the indexes of search engines, and the parsing out of attributes is quick to complete.

BioData Manager

BioData Manager is targeted specifically at Pharmacological and Medical markets. In these industries, the "upstream" raw information from instruments is expensive to capture and its value

can extend beyond its initial use, if all its attributes are preserved and the manageability exists to allow new entities, aggregating new sets of relevance - to be created, as desired.

For many companies, the status quo for extracting data values into the sets they need to support their project (and to comply with FDA archiving regulations) is still error-prone cut-and-paste. BioData Manager automates the discovery, classification, aggregation, and management of the data, easing the pain and expense of FDA reporting. By its flexible nature, it supports more efficient identification, control, and reuse of data.

Partner Products – QStar Data Director

Qstar Data Director, used with *BioData Manager*, provides a file attribute-caching engine that speeds access to the totality of relevant information that you need, even off-line information assets. The Director will e-mail when needed to invoke remounting of offline assets. This allows enterprises to migrate infrequently used data assets to cheaper storage, including tape. For enterprises whose data assets are bulky (medical images), and for research organizations which by their nature can make advance reservations for access to huge quantities of data, Qstar Data Director can make the process more efficient.

Other Uses of Abrevity

Of course, a conspicuous class of instruments to IT folks is the instruments used to manage IT systems. If you are tired of waiting for SIM, and unconvinced of the effectiveness of open movements like Aperi, iBase can be a third way to get what your enterprise needs in a timely fashion. One storage vendor has used Abrevity to deliver data placement dexterity, and others are in discussion. Abrevity believes it could build a specific application for intelligent backup or other IT management functions relatively easily. Future developments can include anything that can be done by managing metadata.

Go to Market Particulars

Abrevity sees itself as the next step beyond traditional databases and executive search. It is not a storage company seeking new capabilities to enhance an array. It offers a ROI tool that starts with industry standard taxonomies and indicates how SLICEBase will work with a sample of customer data, and says its customers

have found no need for professional services to get its products working.

Abrevity is sold as software, not as an appliance. Customers asked for *per terabyte pricing*, rather than *per seat pricing*, because it made it more affordable. This lets customers pay more for files they look inside of and less for others. Since the license is not seat based, an enterprise can give a copy of the software to everyone who needs it. It runs on a laptop, desktop, or server. When more people work on the same data, Abrevity becomes more and more cost-effective.

The basic version is \$499/terabyte. It sees file path metadata, can see inside Microsoft Office and PDF files, and is limited to 3 TB per software instance. For \$5000/terabyte, you get a product that looks inside over 300 kinds of files and scales to petabytes of information. It includes context-based attribute extraction and document summaries.

Conclusion

Abrevity has built its products and set its price to meet the needs of information-centric enterprises. For these enterprises, more information is better. But, even more critical is better *use* of information – something that translates into better and leaner processes – and more routes to more profitability. If your enterprise falls in this category, and most do these days, think about how the Abrevity approach could optimize the way you use information.



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Anne MacFarland is Director of Data Strategies and Information Solutions for The Clipper Group. Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

- ***Reach Anne MacFarland via e-mail at Anne.MacFarland@clipper.com or at 781-235-0085 Ext. 128. (Please dial “128” when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, and “*clipper.com*” are trademarks of The Clipper Group, Inc., and the clipper ship drawings, “*Navigating Information Technology Horizons*”, and “*teraproductivity*” are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.