Bloor

White Paper

Data for everyone: treating data as an asset

"

**Treating data as
a strategic asset
is fundamental to
both data-driven
initiatives and digital
transformation.
It is also essential
to enabling the
democratisation
of data across the
enterprise so that
everyone in the
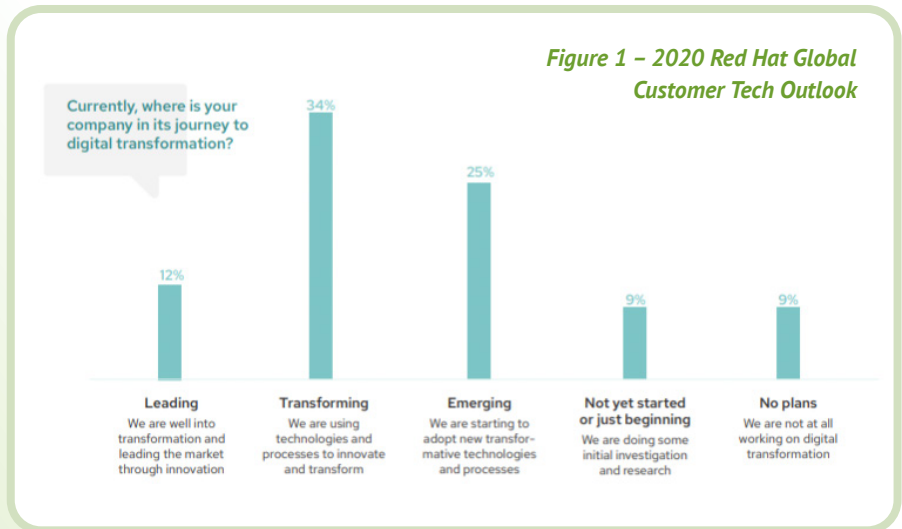organisation can
leverage data
that they trust.**

"

Data for everyone: treating data as an asset

It has become hackneyed to say that data is the new oil. Moreover, given current concerns about climate change, it is a somewhat unfortunate comparison. However, there is no doubt that data – accurate, timely and comprehensive – is essential not just to the day to day running of your business, but also planning for its future. Data is fundamental to digital transformation, which is accelerating. According to research published by Red Hat – see *Figure 1* for the current status – the number of companies having no digital transformation plans has dropped from 32% to 9%, while the number of organisations implementing new technology and processes has increased from 31% to 59%.

Of course, data is not the only contributor to digital transformation, but it is a key asset that can enable those transformations. Moreover, at its heart, digital transformation is about the use of data and automation across the enterprise. And the corollary to this is that the use of your organisation's data becomes pervasive across all of your operations. In effect, data is for everyone.

As witness to this, according to a recently published survey of 1,100 professionals conducted by Insight Avenue on behalf of Experian, *"85% of organisations see data as one of the most valuable assets to their business"* primarily for the reasons illustrated in *Figure 2*.
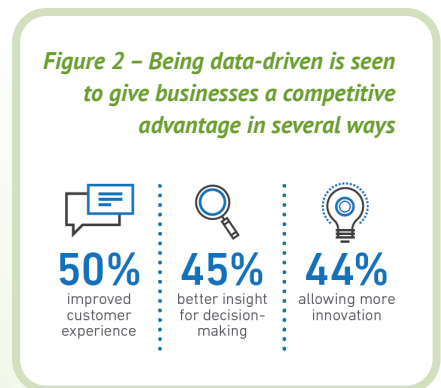
However, there are issues with what Experian calls *"data debt"*, which the company defines as *"the accumulated cost associated with the sub-optimal governance of data assets within the enterprise"*. In other words, the costs of having data that is not complete, timely, accurate and/ or trusted. This is a serious issue that impacts both data-driven initiatives and digital transformation more generally. According to the survey, 78% of companies are affected by this and, in our opinion, a large proportion of the remaining 22% are fooling themselves. Be that as it may, the survey pulls out some interesting statistics with respect to this. To wit:

Figure 1 – 2020 Red Hat Global Customer Tech Outlook

Currently, where is your company in its journey to digital transformation?

- **Leading** 12% — We are well into transformation and leading the market through innovation
- **Transforming** 34% — We are using technologies and processes to innovate and transform
- **Emerging** 25% — We are starting to adopt new transformative technologies and processes
- **Not yet started or just beginning** 9% — We are doing some initial investigation and research
- **No plans** 9% — We are not at all working on digital transformation

- 40% say individuals within the business do not trust data insights
- 35% say they are not able to see the ROI of data management initiatives
- 33% are not able to get value from a new system or technology investment, because of data debt
- 30% of companies are unable to become data-driven

Notwithstanding the above, this paper is concerned with treating data as an asset and we will consider the various aspects of this, one by one. These aspects include having the right data and just the right data, having all the data you need, being able to trust that data, ensuring that it is timely, and being able to share that data to enable collaboration across your environment. Thus, in effect, this paper is about how to avoid or get rid of data debt.

Having completed these discussions, we will discuss how Experian Aperture Data Studio seeks to meet these requirements. We should add that it is coincidental that we have been quoting a survey sponsored by Experian. Had some other vendor or organisation conducted the same or similar research, we would have happily quoted their results, as we would expect them to come up with comparable findings.

Figure 2 – Being data-driven is seen to give businesses a competitive advantage in several ways

50% improved customer experience

45% better insight for decision-making

44% allowing more innovation

Having all the right data

It is self-evident that if you have the wrong data, then any decisions you may make based on that information is likely to be wrong also. But what exactly is the *"right"* data? Put simply, it is the data that you need to do your job. That's all the data you need and no extraneous data that you don't need or that would make your life more difficult in the actual process of doing your job.

Depending on your environment, this is easier said than done. It is relatively trivial if all you need is detail from a CRM or similar system (assuming the data therein is accurate and up to date). But, if you need data from more than one source, things start to get more complicated. The larger your organisation, the more such sources there are likely to be, so getting the *"right"* data becomes increasingly complex. If, for example, you are dealing with customer data, then you want a *"single customer view"* that spans all of your systems that deal with customers. And similarly for products, suppliers, patients, assets and so on. The first issue you have in such environments is simply knowing what data the business holds that is pertinent, and where it resides. You may think you know the answer to these questions but, without some form of enterprise-wide catalogue, there is no way that you can guarantee this.

Even supposing that you are correct in your assumptions about corporate data, there are some fundamental issues. Firstly, do you have access to all of that information? Some of it may be behind firewalls or be data for which you do not have access permission. This may cause a secondary difficulty: how do you get such permissions when you need it? Who do you go to? Organisations should have clear procedures in place to support such demands.

A second issue is if, for example, some of the data is in spreadsheets: are you sure that you are using the most recent version of that spreadsheet? This may be difficult to ascertain if your company does not have spreadsheet governance processes in place. Thirdly, you may need access to third-party data available via the Internet (for example, demographic data). And fourth, all of this data is likely to be different formats: some of it may be in databases (which may themselves have different formats), some of it in text files, some of it in spreadsheets, and so on. In order to work with all of this data you will need to transform the various data sets so that they have a common format. The technology used for this purpose (and related activities) is known as data preparation and it is as useful in operational environments as it is in data science where this technology first emerged.

> "
>
> **The larger your organisation, the more such sources there are likely to be, so getting the *"right"* data becomes increasingly complex.**
>
> "

Trusting the data

Figure 3 –
Results of senior executives surveyed

Figure 3 represents the results of a survey conducted by Varonis. The key figure here is that only just over a third of executives have a high level of trust in the way their enterprise uses data. In other words, 65% do not have that trust. Other such surveys (see Figure 5) have found similar figures with respect to business users in general.

As with wrong and right data, it is self-evident that if you do not trust the data you are working with, then you will feel less confident about any decision you make based on that data than would otherwise be the case. The question therefore becomes: how do you reach a point at which you are happy with the quality of the data with which you are working? And how do you collaborate with your colleagues to improve the quality of your data? We discuss the first of these questions in the remainder of this section. We will consider collaboration in due course.

In practice, enhancing your trust in the data you work with is both a cultural and a technical issue. If you have been in a position where you have been working with untrustworthy data for years, then changing that mind set is going to take time. From a technical perspective, ensuring that data is trustworthy involves a number of elements which are illustrated, courtesy of ScienceSoft, in Figure 4.

We do not need to re-iterate these points, but some comments should be made. One is with respect to uniqueness. It is certainly true that you do not wish to have duplicate records within a single system or database. However, major organisations will frequently have customer records in different systems, as discussed in the previous section. For example, a customer might appear in your CRM system, in your sales order processing system, in your service database, and so on. In this case, you want consistency across the different databases that house these details. To ensure this means having the same ability to identify all relevant data as is required for "right" data. With respect to completeness this should mean that if the service database is updated then relevant information is passed to your CRM and sales systems.

One other point to note is that big data systems, such as data lakes, do not and should not always comply with all of these requirements. However, that does mean that when you pull data out of your big data environment significant data profiling and cleansing of the data may (will) be required. This will typically be done using data preparation, as previously mentioned.
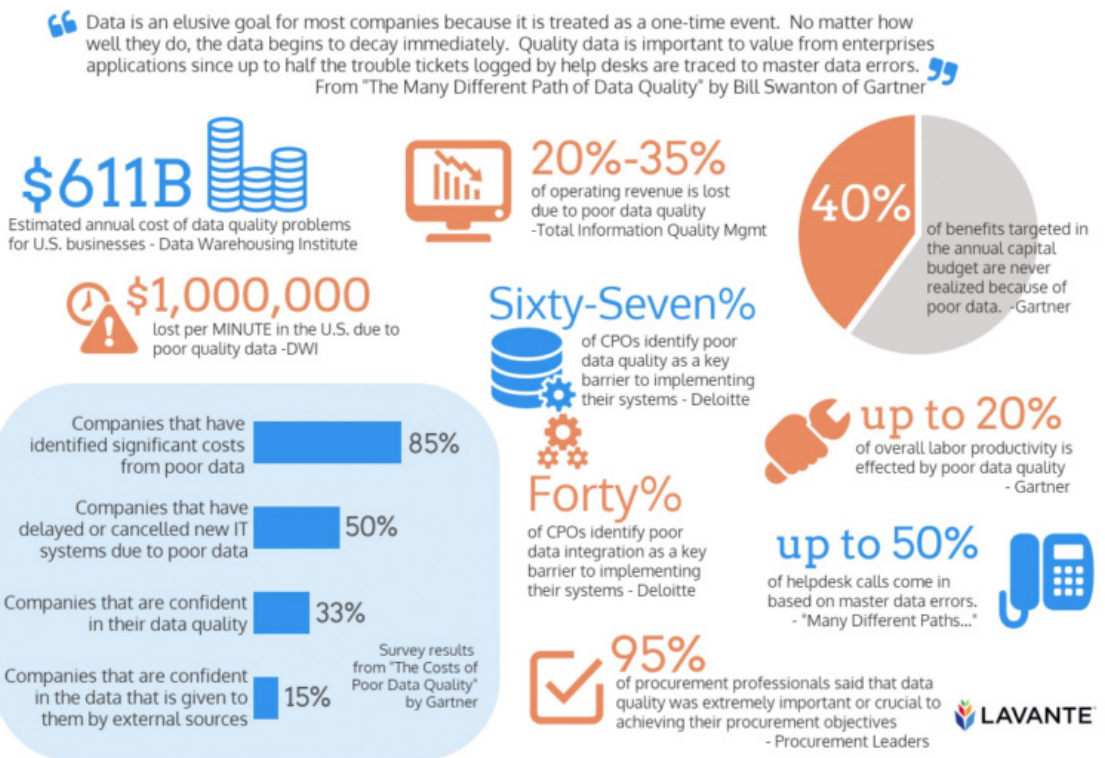

Of Senior Executives Surveyed:

35% have a high level of trust in the way their organization uses data.

62% said technology functions bear responsibility when a machine or algorithm goes wrong.

92% are concerned about the negative impact of data on an organization's reputation.

Figure 4 – Data quality attributes

Attribute	What it means	Example of good practice	Example of bad practice	Metrics
Consistency	No matter where you look in the database, you won't find any contradictions in your data.	Your payment system shows that Jane Brown has made 5 purchases this month, and CRM system contains the same information.	Your payment system shows that Jane Brown has made 5 purchases this month, while CRM system shows she has made only 4.	The number of inconsistencies.
Accuracy	The information your data contains corresponds to reality.	Your customer's name is Jane Brown. And this is exactly how it's reflected in your CRM.	In your CRM, the customer's name is spelled Jane Brawn, though her actual name is Jane Brown.	The ratio of data to errors.
Completeness	All available elements of the data have found their way to the database.	You know that Jane Brown is born on 11/04/1975.	You have no idea how old Jane Brown is, as the date of birth cell is empty.	The number of missing values.
Auditability	Data is accessible and it's possible to trace introduced changes.	You can track down the changes made in Jane's data record. For example, on 12/5/2018, her phone number was changed.	It's impossible to trace down the changes in Jane's record.	% of cells where the metadata about introduced changes is not accessible.
Orderliness	The data entered has the required format and structure.	The entry for December 11, 2018 is in the format 12/11/2018.	The entry for December 11, 2018 is in the format 12/11/18, 12/11/2018 and even 11/12/18 (in your European stores).	The ratio of data of inappropriate format.
Uniqueness	A data record with specific details appears only once in the database.	You have only one record for Jane Brown, born on 11/04/1975, who lives in Seattle.	You have multiple duplicate records for Jane Brown.	The number of duplicates revealed.
Timeliness	Data represents reality within a reasonable period of time or in accordance with corporate standards.	On 02/15/2018, the customer informed you that her name is misspelled in the emails you send her. The customer's name was corrected the next day.	On 02/15/2018, the customer informed you that her name is misspelled in the emails you send her. Her name was corrected only in a month.	Number of records with delayed changes.

The problem with the issue of trusted data is that often the data is actually not trustworthy. Consider *Figure 5*, which is an infographic produced by Levante, quoting a variety of sources. As can be seen, there is a lot of bad data around and it costs a lot of money. While there are plenty of tools available to remediate this issue, the truth is that many companies are still not using them or are not using them holistically across the enterprise. In part this is the fault of the vendors of these products. It is easier for them to sell a solution to a marketing department that recognises that this is an issue, rather than attempt a corporate-wide deployment. And this latter point is often exacerbated by senior executives not being prepared to implement business-wide measures to combat bad data.

" Data is an elusive goal for most companies because it is treated as a one-time event. No matter how well they do, the data begins to decay immediately. Quality data is important to value from enterprises applications since up to half the trouble tickets logged by help desks are traced to master data errors. "
From "The Many Different Path of Data Quality" by Bill Swanton of Gartner

$611B
Estimated annual cost of data quality problems for U.S. businesses - Data Warehousing Institute

$1,000,000
lost per MINUTE in the U.S. due to poor quality data -DWI

Companies that have identified significant costs from poor data — 85%

Companies that have delayed or cancelled new IT systems due to poor data — 50%

Companies that are confident in their data quality — 33%

Companies that are confident in the data that is given to them by external sources — 15%

Survey results from "The Costs of Poor Data Quality" by Gartner

20%-35%
of operating revenue is lost due to poor data quality -Total Information Quality Mgmt

Sixty-Seven%
of CPOs identify poor data quality as a key barrier to implementing their systems - Deloitte

Forty%
of CPOs identify poor data integration as a key barrier to implementing their systems - Deloitte

95%
of procurement professionals said that data quality was extremely important or crucial to achieving their procurement objectives - Procurement Leaders

40%
of benefits targeted in the annual capital budget are never realized because of poor data. -Gartner

up to 20%
of overall labor productivity is effected by poor data quality - Gartner

up to 50%
of helpdesk calls come in based on master data errors. - "Many Different Paths..."

LAVANTE

The right time

Timeliness was highlighted in the last section, but it merits a discussion of its own. Once again, it should be self-evident that you want to work with data that is current, not data that is out-of-date. However, this is an issue, as *Figure 5* points out, because *"data begins to decay immediately"*. Moreover, this is a continuous process. You should assume, from the point of view of data trust that data quality is always decaying. You don't need, therefore, to know when it starts to decay, it is ever present. As a result data quality remediation needs to be treated as an ongoing process and not a one-time event.

The reason for this is straightforward. *Figure 6*, from Radius Business Graph shows just how many business contact details will change over a typical three-month period. Even if we assume that half of the figures given represent the same event (a business moving address may well change its phone number), that still represents more than a 10% quarterly churn rate. More than 40% per annum.

Of course, it is true that both consumers and businesses may fail to tell you when they move. However, they will notify tax authorities, phone companies, postal services and so on and there are third party providers that will collect this information and can make it available to you. It should be a matter of routine to compare such data with your in-house information on a regular basis, so that your databases can be kept up-to-date.

> **"**
> **You should assume, from the point of view of data trust that data quality is always decaying.**
> **"**

Figure 6 –
Data decays significantly
after only 3 months

Based on average data rates, after only 3 months a database with 10,000 business will have the following data outdated:

210 Phone Numbers Will Disconnect
Businesses mover or close

418 Phone Numbers Will Change
Phone numbers are updated or reassigned between businesses

592 Business Addresses Will Become Invalid
Businesses move or go out of business

757 Contacts Won't Connect
Contacts move to other jobs or are no longer reachable

252 Emails Become Invalid
Emails become invalid after contacts move or businesses cease operations

Source: Radius Business Graph℠, August 2016.

Collaboration

According to a study published by the Institute for Corporate Productivity, *"companies promoting collaborative working are five times as likely to be high performing (compared to those that do not)"*. Of course, collaborative working applies to many areas of business life and not just to collaboration with respect to data assets. And collaboration is primarily a cultural issue though it can be assisted by relevant software tools that have features specifically designed to enable collaboration. As far as the current discussion goes there are a number of features of relevant products that can help to enable collaboration.

To begin with, decisions are often made in collaboration rather than by a single person. Enabling this means being able to share the data upon which any such decision rests. However, it is important than everyone involved in any decision is seeing the same data as everyone else. There is a danger, especially with spreadsheets but also with editable reports or documents, that users will simply email copies of that data to those involved, who may in turn create new versions of those assets. Best practice would involve either making this impossible – perhaps by only making the data visible via a portal – or by implementing version control that will enable surety about everyone seeing the same information.

Another element of collaboration is with respect to *"right"* data and enterprise data sources. It is often the case that colleagues are working with complementary and/or overlapping sets of data. A useful feature is to be able to *"like"* particular datasets, or give them star ratings, indicating that these sources hold particularly or not particularly useful information. This can help to speed up the process of finding the *"right"* data. Software that recommend particular datasets (perhaps based on built-in machine learning) will also be useful in this respect.

Finally, there is a third aspect of collaboration that relates to business users working with technical experts. This is typically accomplished by means of *"personas"*, where different user communities, depending on their organisational role, access the data they are working with via this persona. This allows business users, data stewards, data scientists, technical experts and so on, all access the same data, via the same portal, but with a view on that data that is specific to the requirements of that persona. Such a portal will also support the attachment of notes and comments that can be shared both within and across personas.

> *...companies promoting collaborative working are five times as likely to be high performing (compared to those that do not).*

Experian Aperture Data Studio

In this paper we have outlined why it is important to treat data as a strategic asset, along with the things you need to think about in order to treat it as such. Experian is one company that specialises in helping businesses to achieve this, and its primary tool for this purpose is Experian Aperture Data Studio.

Experian Aperture Data Studio is a self-service data quality and enrichment platform that provides in-depth data profiling and data quality analysis. Along with the provision of globally curated data sets, it is intended to support the creation of consistent, accurate and holistic views of customers (or patients, clients, criminals and so on). You can also accompany this data analysis activity with data preparation and manipulation pipelines to support data quality improvement.

The product supports all the sorts of functional requirements that are shown in *Figure 7*. The product has been designed to work within existing technology stacks. Thus, for example, the product's data preparation capabilities which are largely targeted at operational rather than analytic environments, are complementary to both data cataloguing and data governance solutions. The integration with data catalogues means that, provided these are enterprise-wide, then you should be able to identify all the data you need, and then using Experian's data preparation capabilities you can transform and join this data so that you end up with the *"right"* data.

As far as trusting the data is concerned, this is Experian Aperture Data Studio's bread and butter. All of the sorts of data quality attributes identified in *Figure 4* are provided by the platform and we don't need to dwell on these capabilities. However, with respect to timeliness, we should add that Experian provides a range of other data quality capabilities that can be used seamlessly with Aperture Data Studio to allow you to periodically check that things such as address details, and telephone numbers and emails, are current and up-to-date.

A significant business challenge to solve is the identification and resolution of duplicate records. Experian Aperture Data Studio's Find Duplicates matching engine combines standard and customisable data matching rules along with fuzzy-matching

algorithms to converge on the best possible consumer records. When this capability is accompanied by region-specific Experian *'pinning'* data (see *Figure 8*) that is based on current and historical contact data, companies are able to converge more accurately on known persons and households.

The ability of Aperture Data Studio to find duplicate records is principally concerned with identifying and resolving duplicate records in contact data. As such, Data Studio comes with pre-prepared data tagging data fingerprints which allow for automatic machine learning data classification. In addition, a machine learning and artificial intelligence-based training and rule optimisation engine enables you to create your own data fingerprints and extends the applicability of the platform to data classification beyond just customer contact data.

Finally, with respect to collaboration, Experian Aperture Data Studio is persona-based (domain experts, data stewards, business analysts and so forth) and, in the latest release there is a strong emphasis on supporting collaboration between the people who need to use your data for operational and analytic reasons and those that have a more technical relationship with and understanding of that data.

> **As far as trusting the data is concerned, this is Experian Aperture Data Studio's bread and butter.**

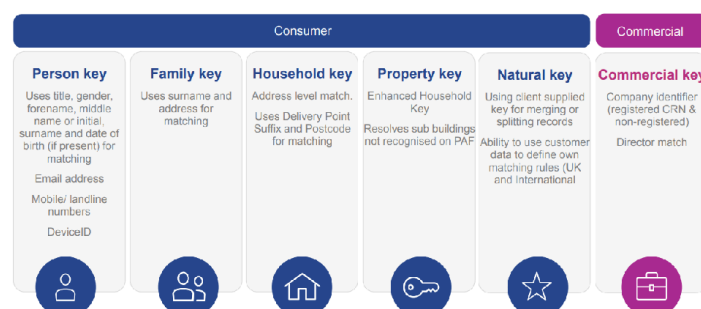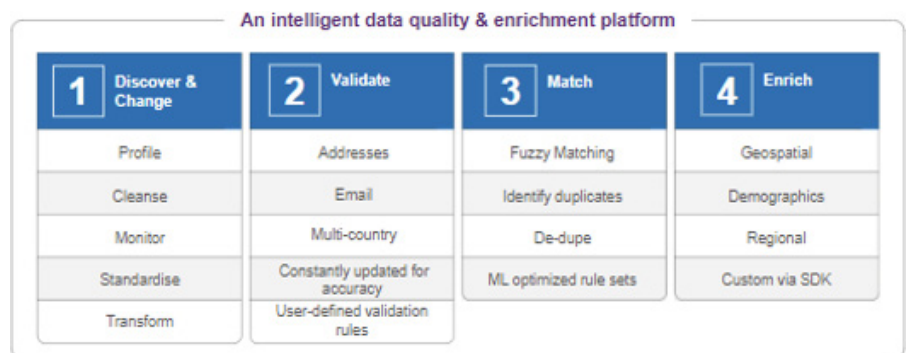

Figure 7 – Aperture Data Studio

Figure 8 – Experian 'pinning' data

Conclusion

T reating data as a strategic asset is fundamental to both data-driven initiatives and digital transformation. It is also essential to enabling the democratisation of data across the enterprise so that everyone in the organisation can leverage data that they trust. Unfortunately, it is a sad indictment of the state of corporate awareness that Bloor Research first published a report on this subject in the last century and that we are still writing about it as a major issue more than twenty years later. Be that as it may, in this paper we have outlined, without going into too much technical detail, why the treatment of data as an asset is important, how it enables the use of data for everyone, and the primary things you should think about when adopting that approach. We have also briefly described how Experian Aperture Data Studio can meet the objectives of treating data as an asset and in supporting data democratisation.

"

...it is a sad indictment of the state of corporate awareness that Bloor Research first published a report on this subject in the last century and that we are still writing about it as a major issue more than twenty years later.

"

FURTHER INFORMATION
Further information about this subject is available from
www.bloorresearch.com/update/2567

About the author
PHILIP HOWARD
Research Director / Information Management

P hilip started in the computer industry way back in 1973 and has variously worked as a systems analyst, programmer and salesperson, as well as in marketing and product management, for a variety of companies including GEC Marconi, GPT, Philips Data Systems, Raytheon and NCR.

After a quarter of a century of not being his own boss Philip set up his own company in 1992 and his first client was Bloor Research (then ButlerBloor), with Philip working for the company as an associate analyst. His relationship with Bloor Research has continued since that time and he is now Research Director, focused on Information Management.

Information management includes anything that refers to the management, movement, governance and storage of data, as well as access to and analysis of that data. It involves diverse technologies that include (but are not limited to) databases and data warehousing, data integration, data quality, master data management, data governance, data migration, metadata management, and data preparation and analytics.

In addition to the numerous reports Philip has written on behalf of Bloor Research, Philip also contributes regularly to *IT-Director.com* and *IT-Analysis.com* and was previously editor of both *Application Development News* and *Operating System News* on behalf of Cambridge Market Intelligence (CMI). He has also contributed to various magazines and written a number of reports published by companies such as CMI and The Financial Times. Philip speaks regularly at conferences and other events throughout Europe and North America.

Away from work, Philip's primary leisure activities are canal boats, skiing, playing Bridge (at which he is a Life Master), and dining out.

Bloor overview

Technology is enabling rapid business evolution. The opportunities are immense but if you do not adapt then you will not survive. So in the age of Mutable business Evolution is Essential to your success.

We'll show you the future and help you deliver it.

Bloor brings fresh technological thinking to help you navigate complex business situations, converting challenges into new opportunities for real growth, profitability and impact.

We provide actionable strategic insight through our innovative independent technology research, advisory and consulting services. We assist companies throughout their transformation journeys to stay relevant, bringing fresh thinking to complex business situations and turning challenges into new opportunities for real growth and profitability.

For over 25 years, Bloor has assisted companies to intelligently evolve: by embracing technology to adjust their strategies and achieve the best possible outcomes. At Bloor, we will help you challenge assumptions to consistently improve and succeed.

Copyright and disclaimer