2017 Edition

End User Data Preparation

Wisdom of Crowds® Series

Licensed to Datawatch
Disclaimer:

This report is for informational purposes only. You should make vendor and product selections based on multiple information sources, face-to-face meetings, customer reference checking, product demonstrations, and proof-of-concept applications.

The information contained in this Wisdom of Crowds® market study report is a summary of the opinions expressed in the online responses of individuals that chose to respond to our online questionnaire and does not represent a scientific sampling of any kind. Dresner Advisory Services, LLC shall not be liable for the content of this report, the study results, or for any damages incurred or alleged to be incurred by any of the companies included in the report as a result of the report’s content.

Reproduction and distribution of this publication in any form without prior written permission is forbidden.
Definitions

**Business Intelligence Defined**
Business intelligence (BI) is “knowledge gained through the access and analysis of business information.”

Business Intelligence tools and technologies include query and reporting, OLAP (online analytical processing), data mining and advanced analytics, end-user tools for ad hoc query and analysis, and dashboards for performance monitoring.


**End User Data Preparation Defined**
End User Data Preparation is a "self-service" capability for end users to model, prepare, and combine data prior to analysis. This may complement traditional IT-driven Data Quality/ETL processes or may be used independently.
Contents
Definitions .................................................................................................................................................. 3
  Business Intelligence Defined .................................................................................................................. 3
  End User Data Preparation Defined ..................................................................................................... 3
Introduction ............................................................................................................................................... 6
About Howard Dresner and Dresner Advisory Services ................................................................. 7
About Jim Ericson .................................................................................................................................... 8
Findings and Analysis ............................................................................................................................. 9
Focus of Research .................................................................................................................................... 9
Benefits of the Study .............................................................................................................................. 10
  Consumer Guide .................................................................................................................................. 10
  Supplier Tool ......................................................................................................................................... 10
  External Awareness ............................................................................................................................... 10
  Internal Planning ................................................................................................................................ 10
Survey Method and Data Collection ...................................................................................................... 11
  Data Quality ......................................................................................................................................... 11
Executive Summary .............................................................................................................................. 12
Study Demographics ............................................................................................................................... 13
  Geography ........................................................................................................................................... 13
  Functions .............................................................................................................................................. 14
  Vertical Industries ................................................................................................................................. 15
  Organization Size ................................................................................................................................ 16
Analysis of Findings ................................................................................................................................. 17
  Importance of End-User Data Preparation ............................................................................................ 18
  Effectiveness of Current Approach to End-User Data Preparation .................................................... 25
  Frequency of End-User Data Preparation ............................................................................................. 31
  Frequency of End-User Data Preparation Enrichment with Third-Party Data ...................................... 37
  End-User Data Preparation Usability Features .................................................................................... 43
  End-User Data Preparation Data Integration Features ......................................................................... 49
  End-User Data Preparation Manipulation Features ............................................................................... 55
  End-User Data Preparation Supported Outputs .................................................................................... 61
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-User Data Preparation Deployment Features</td>
<td>66</td>
</tr>
<tr>
<td>Location of End-User Data Preparation Capabilities</td>
<td>72</td>
</tr>
<tr>
<td>End-User Data Preparation: Standalone versus Inclusion with Other Software</td>
<td>77</td>
</tr>
<tr>
<td>Industry Support for End-User Data Preparation</td>
<td>83</td>
</tr>
<tr>
<td>Industry Support for End-User Data Preparation Usability</td>
<td>84</td>
</tr>
<tr>
<td>Industry Support for End-User Data Preparation Integration</td>
<td>85</td>
</tr>
<tr>
<td>Industry Support for End-User Data Preparation Output Options</td>
<td>86</td>
</tr>
<tr>
<td>Industry Support for End-User Data Preparation Data Manipulation Features</td>
<td>87</td>
</tr>
<tr>
<td>Industry Support for End-User Data Preparation Deployment Features</td>
<td>88</td>
</tr>
<tr>
<td>Industry Support for End-User Data Preparation—Cloud versus On-Premises</td>
<td>89</td>
</tr>
<tr>
<td>End-User Data Preparation Vendor Ratings</td>
<td>90</td>
</tr>
<tr>
<td>Other Dresner Advisory Services Research Reports</td>
<td>91</td>
</tr>
<tr>
<td>Appendix: End User Data Preparation Survey Instrument</td>
<td>92</td>
</tr>
</tbody>
</table>
Introduction
This year we celebrate the tenth anniversary of Dresner Advisory Services! Our thanks to all of you for your continued support and ongoing encouragement.

Since our founding in 2007, we have worked hard to set the “bar” high—challenging ourselves to innovate and lead the market—offering ever greater value with each successive year.

Our first market report in 2010 set the stage for where we are today. Since that time, we have expanded our agenda and have added new research topics every year. For 2017, we plan to release 15 major reports, including our original BI flagship report—in its eighth year of publication!

This publication marks our third annual End User Data Preparation report. End user data preparation is a topic that resonates strongly with organizations—and especially with power users and analysts that have been relegated to using whatever tools were available for the purpose—regardless of limitations.

An important step towards the ongoing trend of user empowerment and self-service business intelligence, end user data preparation is driving an increasing amount of investment on both demand and supply sides of the equation.

For 2017, we added additional criteria—“raising the bar” for what is required as a part of an end user data preparation solution.

We hope you enjoy this report!

Best,

Chief Research Officer
Dresner Advisory Services
About Howard Dresner and Dresner Advisory Services

The DAS End User Data Preparation Market Study was conceived, designed and executed by Dresner Advisory Services, LLC—an independent advisory firm—and Howard Dresner, its President, Founder and Chief Research Officer.

Howard Dresner is one of the foremost thought leaders in business intelligence and performance management, having coined the term “Business Intelligence” in 1989. He has published two books on the subject, *The Performance Management Revolution – Business Results through Insight and Action* (John Wiley & Sons, Nov. 2007) and *Profiles in Performance – Business Intelligence Journeys and the Roadmap for Change* (John Wiley & Sons, Nov. 2009). He lectures at forums around the world and is often cited by the business and trade press.

Prior to Dresner Advisory Services, Howard served as chief strategy officer at Hyperion Solutions and was a research fellow at Gartner, where he led its business intelligence research practice for 13 years.

Howard has conducted and directed numerous in-depth primary research studies over the past two decades and is an expert in analyzing these markets.

Through the Wisdom of Crowds® Business Intelligence market research reports, we engage with a global community to redefine how research is created and shared. Other research reports include:

- **Wisdom of Crowds “Flagship” Business Intelligence Market study**
- **Advanced and Predictive Analytics**
- **Analytical Data Infrastructure**
- **Big Data Analytics**
- **Cloud Computing and Business Intelligence**
- **Collective Insights®**
- **Internet of Things and Business Intelligence**
- **Location Intelligence**
- **Natural Language Analytics**

Howard conducts a weekly Twitter “tweetchat” on Fridays at 1:00 p.m. ET. During these live events the #BIWisdom “tribe” discusses a wide range of business intelligence topics.

You can find more information about Dresner Advisory Services at www.dresneradvisory.com.
About Jim Ericson

Jim Ericson is a research director with Dresner Advisory Services.

Jim has served as a consultant and journalist who studies end-user management practices and industry trending in the data and information management fields.

From 2004 to 2013 he was the editorial director at Information Management magazine (formerly DM Review), where he created architectures for user and industry coverage for hundreds of contributors across the breadth of the data and information management industry.

As lead writer he interviewed and profiled more than 100 CIOs, CTOs, and program directors in a 2010-2012 program called “25 Top Information Managers.” His related feature articles earned ASBPE national bronze and multiple Mid-Atlantic region gold and silver awards for Technical Article and for Case History feature writing.

A panelist, interviewer, blogger, community liaison, conference co-chair, and speaker in the data-management community, he also sponsored and co-hosted a weekly podcast in continuous production for more than five years.

Jim’s earlier background as senior morning news producer at NBC/Mutual Radio Networks and as managing editor of MSNBC’s first Washington, D.C. online news bureau cemented his understanding of fact-finding, topical reporting, and serving broad audiences.
Findings and Analysis
In this report, we present the deliverables for our End User Data Preparation Market Study based upon data collection from July through October 2016.

Focus of Research
In this study, we address key end-user data preparation issues including:

- Perceptions and intentions surrounding end-user data preparation
- End-user requirements and features:
  - Usability features
  - Integration features
  - Manipulation features
  - Output options
  - Deployment options
- Industry support for end-user data preparation
- User requirements versus industry capabilities
- Vendor ratings
Benefits of the Study
This DAS End User Data Preparation Market Study provides a wealth of information and analysis, offering value to both consumers and producers of business intelligence technology and services.

Consumer Guide
As an objective source of industry research, consumers use the DAS End User Data Preparation Market Study to understand how their peers leverage and invest in end-user data preparation and related technologies.

Using our unique vendor performance measurement system, users glean key insights into BI software supplier performance, which enables:

- Comparisons of current vendor performance to industry norms
- Identification and selection of new vendors

Supplier Tool
Vendor licensees use the DAS End User Data Preparation Market Study in several important ways:

External Awareness
- Build awareness for business intelligence markets and supplier brands, citing the DAS End User Data Preparation Market Study trends and vendor performance
- Gain lead and demand generation for supplier offerings through association with the DAS End User Data Preparation Market Study brand, findings, webinars, etc.

Internal Planning
- Refine internal product plans and align with market priorities and realities as identified in the DAS End User Data Preparation Market Study
- Better understand customer priorities, concerns, and issues
- Identify competitive pressures and opportunities
Survey Method and Data Collection
As with all of our Wisdom of Crowds® Market Studies, we constructed a survey instrument to collect data and used social media and crowdsourcing techniques to recruit participants.

Data Quality
We carefully scrutinized and verified all respondent entries to ensure that only qualified participants were included in the study.
Executive Summary

- End-user data preparation ranks 15th in importance among 30 BI topics under study in 2017 (p. 18).
- Two thirds of respondents say end-user data prep is "critical or "very important," and sentiment over time has remained consistent (pp. 19-24). Industry support remains consistently high over time (p. 83).
- A large majority of organizations say their current end-user data preparation approach is "highly effective" or "somewhat effective." Insurance and Financial Services respondents are the most confident (pp. 25-30).
- Two-thirds of respondents "constantly" or "frequently" make use of end-user data preparation and have increased use over time. Finance and Marketing/Sales are the greatest users; Consumer Products is the most involved industry (pp. 31-36).
- A majority of organizations enrich end-user data preparation with third-party data, though third-party data use has not accelerated greatly over time and is not a "front-burner" priority. Marketing/Sales are the most likely users. (pp. 37-42).
- Respondents are interested in a full range of usability features for data prep, led by "immediate preview/feedback" (pp. 43-48). Industry support for usability is good to strong (p. 84).
- Demand for end-user data prep integration features is strong, steady, and led by conventional integrations with flat files, databases, and joins/merges (pp. 49-54). Interest in big data integration is lowest in North America. Industry support for user integration needs is very robust (p. 85).
- Among end-user data prep manipulation features, the "ability to aggregate and group data" and "ability to pivot data "are most critical to users (pp. 55-60). The industry supports manipulation features well (p. 87).
- The most important data prep tool outputs are to flat files formats, outputs to databases, and direct to business intelligence tools (pp. 61-65). The industry supports current user needs along with newer formats (p. 88).
- The most important data prep deployment features are scheduling/reviewing transformations and the ability to iteratively sample data (pp. 66-71). Industry support for deployment features is mostly good but not entirely aligned (p. 88).
- Users prefer on-premises data prep deployment to private or public cloud (pp. 72-76). The vendor industry fully supports user preferences for different deployment locations (p. 89).
- Users feel strongly that data prep tools should be included in BI tools (as opposed to standalone) for a seamless interaction, a sentiment that has grown over time (pp. 77-82).
Study Demographics
Our sample includes a cross-section of data across geographies, functions, organization sizes, and vertical industries. We believe that, unlike other industry research, we offer a more characteristic sample and better indicator of true market dynamics.

Geography
Survey respondents represent a mix of global geographies. Sixty-four percent represent North America (including five Canadian provinces and a majority of U.S. states). Twenty-five percent work in EMEA, 7 percent in Asia Pacific, and 4 percent in Latin America (fig. 1).

Figure 1 – Geographies represented
Functions

Information Technology accounts for the largest group of respondents (31 percent) by function. About 24 percent come from the Business Intelligence Competency Center (BICC). Executive Management and Finance are the next most represented (fig. 2).

Tabulating results by function enables us to compare and contrast the plans and priorities of different departments within organizations.

Figure 2 – Functions represented
Vertical Industries
Survey participants represent a wide range of vertical industries led by Technology (12 percent), Healthcare (11 percent), Financial Services (10 percent), and consulting (fig.3). We allow and encourage the participation of consultants, who often have deeper industry knowledge than their customer counterparts. Third-party relationships give us insight into the partner ecosystem for BI vendors.

Figure 3 – Vertical industries represented
**Organization Size**

Our survey sample includes a mix of small, medium, and large organizations (fig. 4). In 2017, small organizations (1-100 employees) account for 28 percent of the sample, and mid-sized organizations (101-1,001 employees) account for 26 percent of the sample. Large organizations (>1,000 employees) account for the remaining 47 percent, with very large organizations (>5,000 employees) accounting for 23 percent.

Segmenting respondents by organization size helps us identify differences in behavior, attitudes, and planning often related to headcount.

![Organization Sizes Represented](image_url)

*Figure 4 – Organization sizes represented*
Analysis of Findings
In 2017, (our third annual) End User Data Preparation Market Study, we examine the nature of end-user data preparation, exploring user sentiment and perceptions, the nature of current implementations, and plans for the future.
Importance of End-User Data Preparation

Among technologies and initiatives strategic to business intelligence in 2017, end-user data preparation (aka blending) ranks 15th, at the midpoint of 30 topics we currently study (fig. 5). Thus, end-user data preparation importance trails traditional topics including reporting, dashboards, end-user self-service, data visualization, and data discovery. But it is well ahead of many familiar topics including cloud computing, big data, and the Internet of Things. We believe the relative strategic importance users attach to end-user data preparation underscores the value attached to end-user empowerment and self-service generally.

Figure 5 – Technologies and initiatives strategic to business intelligence
In this, our third year of focused study of end-user data preparation, respondents’ perceived importance of end-user data preparation is very high and in line with user demands for self-service business intelligence and user autonomy (fig. 6). Sixty-seven percent of all respondents say end-user data preparation is either “critical” or “very important.” About 88 percent of respondents say end-user data preparation is, at minimum, “important.” Just 3 percent say end-user data preparation is “not important.”

![Importance of End-User Data Preparation](image)

Figure 6 – Importance of end-user data preparation
Across three years of data, the perceived importance of end-user data preparation remains largely consistent (fig. 7). In 2017, mean level importance stands at 3.86, a score approaching "very important;" and, increasingly, respondents view data prep as an expected component within a business intelligence tool. Seventy-seven percent of respondents now say end-user data preparation is either "critical" or "very important." Only 13 percent say the topic is "somewhat important" or "not important."

![Importance of End-User Data Preparation 2015-2017](image)

*Figure 7 – Importance of end-user data preparation 2015-2017*
Among the functions we sampled in 2017, Finance and the BICC report the highest estimation of the “critical” and overall importance of end-user data preparation (fig. 8). That said, mean levels of importance are consistently high across functions with mean importance of 3.7 to 4.0 (“very important”). Favorability among large majorities of users across functions indicates that end-user data preparation is critical to front-end processes related to revenue and market share.

![Importance of End-User Data Preparation by Function](image)

*Figure 8 – Importance of end-user data preparation by function*
By geography, respondents in North America, Asia Pacific, and EMEA have consistently high opinions of the importance of end-user data preparation (fig. 9). Asia-Pacific respondents are most likely (47 percent) to consider end-user data preparation "critical." Asia-Pacific respondents also report a greater diversity of opinion.

Figure 9 – Importance of end-user data preparation by geography
The importance of end-user data preparation extends across organizations of different sizes (fig. 10). Small organizations of 1-100 employees are most likely to consider end-user data preparation "critical" (38 percent) or "very important" (35 percent). Nearly 70 percent of organizations with more than 1,000 employees say the technology is "very important" or "important." Sentiment is slightly lower at mid-sized organizations (101-1,000 employees), where 60 percent consider the technology "very important" or "critical."

**Figure 10 – Importance of end-user data preparation by organization size**

![Importance of End-User Data Preparation by Organization Size](chart.png)
Mean perceived importance of end-user data preparation is more variable by industry than by other measures (fig. 11). In our 2017 sample, Energy respondents attach the most "critical" importance (70 percent), compared to other industries. Automotive and Education, followed by Insurance and Manufacturing, report the next highest scores, near or above "very important." Other industries score end-user data prep as between "important" and "very important."

![Importance of End-User Data Preparation by Industry](image-url)

Figure 11 – Importance of end-user data preparation by industry
**Effectiveness of Current Approach to End-User Data Preparation**

In 2017, a large majority of organizations say their current end-user data preparation approach is "highly effective" or "somewhat effective" (fig. 12). Just 4 percent say their current approach is "totally ineffective." These results imply good levels of interaction and experience with end-user data preparation, likely in the context of self-service and user autonomy, which are prime drivers for data preparation and business intelligence generally.

![Current Approach to End-User Data Preparation](image)

*Figure 12 – Current approach to end-user data preparation*
Across three years of data collection, respondents say their current approach to end user data preparation has improved steadily over time (fig. 13). The number that report "highly effective" use reaches 18 percent in 2017, while "somewhat effective" organizations reach 56 percent, both all-time highs. In the same time period, fewer organizations report "somewhat ineffective" or "totally ineffective" use of the technology. Such a finding implies maturity and better/more effective penetration of end-user data prep to involved parties.

![Current Approach to End-User Data Preparation 2015-2017](image)

*Figure 13 – Current approach to end-user data preparation 2015-2017*
All functions report "somewhat effective" mean levels of satisfaction with end-user data preparation (fig. 14). In 2017, respondents from Executive Management and the BICC are most likely to see their data preparation as effective. Executive Management tends to "look down the mountain" at perceived effectiveness, which likely has more relevance in the appraisals of Finance or Marketing/Sales. In that regard, Marketing/Sales is least satisfied, while Finance appears to do the best job among hands-on users.

Figure 14 – Current approach to end-user data preparation by function
There is no pronounced geographical/regional difference in the perceived effectiveness of end-user data preparation (fig. 15). Latin American and EMEA respondents are most likely to consider their current approach to end-user data preparation "somewhat effective" or "highly effective." More critical users are found in Asia Pacific and North America, though no region is more than 10 percent likely to say their efforts are "totally ineffective."

![Current Approach to End-User Data Preparation by Geography](image)

**Figure 15 – Current approach to end-user data preparation by geography**
Organizations of different sizes report rather consistent mean level views of their effectiveness with the use of end-user data preparation (fig. 16). While positive sentiment is slightly lower in small (1-100 employees) and mid-sized (101-1,000 employees) organizations, all organizations we sampled in 2017 generally consider their current approach "somewhat effective."

![Current Approach to End-User Data Preparation by Organization Size](image-url)

*Figure 16 – Current approach to end-user data preparation by organization size*
Perceived effectiveness of end-user data preparation varies more by industry than by other dimensions (fig. 17). Insurance organizations (with extensive back-office users in actuary and underwriting roles), unanimously agree their efforts are either "somewhat effective" or "highly effective." Financial Services is likewise confident in end-user data preparation success, while Healthcare and Education consider themselves less effective. Again, overall industry impressions are that end-user data preparation efforts are in the range of "somewhat effective."

Figure 17 – Current approach to end-user data preparation by industry
**Frequency of End-User Data Preparation**

Sixty-seven percent of respondents say they "constantly" or "frequently" make use of end-user data preparation (fig. 18). We cannot distinguish whether end-user efforts are one-off or regular practice, but overall usage of end-user data preparation is high. Twenty-nine percent say they only "occasionally" require end-user data preparation; the remaining 6 percent "rarely" or "never" do.

![Frequency of End-User Data Preparation](image-url)

*Figure 18 – Frequency of end-user data preparation*
Across three years of data collection, respondents report increasing use of end-user data preparation (fig. 19). In 2017, "constant" and "occasional" use increased, "frequent" usage is flat, while "rare" and "never" use declined. At a high level, we believe there is more activity, whether among a static group of users, an increasing audience, or both.

Figure 19 – Frequency of end-user data preparation 2015-2017
We would expect that issues of business performance and revenue would drive the frequency of use of end-user data preparation. In 2017, Finance and Marketing/Sales are the greatest users among identified roles (fig. 20). Activity levels begin to decrease in the Business Intelligence Competency Center, R&D, IT, and Project/Program Managers, perhaps indicating that users of data prep are self-empowered more than they are supported by services. We note that Marketing/Sales are more frequently involved with data prep; yet, they are dissatisfied with its effectiveness (fig. 14, p. 27).

Figure 20 – Frequency of end-user data preparation by function
The frequency of constant end-user data preparation use is mostly consistent across geographies. Ninety percent or more of respondents in all geographies are, at minimum, occasional users (fig. 21). The fewest "constant" users are in EMEA. Overall, frequency measurements by mean are between "frequent" and "constant" across all geographies.

Figure 21 – Frequency of end-user data preparation by geography
Mean frequency of end-user data preparation is very consistent across organizations of different sizes (fig. 22). "Constant" use declines slightly as organization size increases, and to a greater extent at organizations with more than 5,000 employees. Combined "constant" and "frequent" use is nonetheless greatest at very large organizations.

Figure 22 – Frequency of end-user data preparation by organization size
End-user data preparation frequency varies somewhat by industry (fig. 23). As we might expect, consumer products and automotive respondents with broad stocks of SKUs and inventory are the most frequent users in our 2017 sample. All industries report levels of activity above 3.5, which informs us of a great deal of activity across industries as well as other measures.

Figure 23 – Frequency of end-user data preparation by industry
Frequency of End-User Data Preparation Enrichment with Third-Party Data

A majority of organizations "constantly," "frequently," or "occasionally" enrich end-user data preparation with third-party data (fig. 24). Still, just 6 percent are "constant" users of non-proprietary data. Overall, we see a fairly broad spectrum of third-party data use: 22 percent are "frequent" users while another 22 percent "rarely" use third-party data.

![Frequency of End-User Data Preparation Enrichment with Third-Party Data](image_url)

Figure 24 – Frequency of end-user data preparation enrichment with third-party data
Across three years of data, respondents give a mixed view of third-party data use in conjunction with end-user data preparation (fig. 25). Most notably, "constant" users of external data decreased over time in favor of "occasional" users. We do not see a radical uptake in third-party data use, and organizations appear to be grappling mostly with internal data. If we were to apply a mean score to third-party data use over time, we would find it mostly flat.

Figure 25 – Frequency of end-user data preparation enrichment with third-party data 2015-2017
By function, Marketing/Sales is the most likely to constantly or frequently enrich end-user data preparation with third-party data (fig. 26). This is consistent with contextual use of business intelligence alongside credit, mapping, social media, and consumer profiling. As we would expect, Finance is the least interested in third-party data enrichment, and lower levels of use in IT and R&D appear to show that the topic is not a front-burner priority in 2017.

![Frequency of End-User Data Preparation Enrichment with Third-Party Data by Function](image)

*Figure 26 – Frequency of end-user data preparation enrichment with third-party data by function*
Interest in third-party data enrichment of end-user data preparation is fairly consistent across geographies (fig. 27). Respondents in Latin America are most likely to use third-party sources; other regions report less but more consistent usage patterns. Mean sentiment toward third-party data use is generally near the 3.0 or "occasional" level of use across regions.

![Frequency of End-User Data Preparation Enrichment with Third-Party Data by Geography](image-url)

Figure 27 – Frequency of end-user data preparation enrichment with third-party data by geography
The use of third-party data enrichment in end-user data preparation is largely consistent across organizations of different sizes (fig. 28). Very large organizations (>5,000 employees) are the most likely "constant" users while mid-sized organizations (101-1,000 employees) account for the most frequent users. Between 30 and 40 percent of organizations of any size "rarely" or "never" use third-party sources.

Figure 28 – Frequency of end-user data preparation enrichment with third-party data by organization size
We would expect industries that are highly transactional or sensitive to customer attitudes, churn, and loyalty to be frequent users of third-party data enrichment. In this regard, we are not surprised to see Consumer Products, Insurance, and Telecommunications atop this measurement (fig. 29). From an overall high of 4.0 (frequent) usage, third-party data enrichment thereafter drops rather precipitously in Healthcare, Financial Services, and other industries, though prospects may be in the offing as more data sources come online.

Figure 29 – Frequency of end-user data preparation enrichment with third-party data by industry
End-User Data Preparation Usability Features

Respondents have high interest in a full range of end-user data preparation usability features, all of which they consider "important" to "very important" (fig. 30). We believe this reflects good understanding of needs and high expectations for data preparation features associated with BI/analytics usage. A feature we added for 2017, "immediate preview and feedback," debuted as a top requirement and is at least "very important" to almost 80 percent of respondents. "Visual user interface" and "technical expertise not required" are also very important to large majorities of respondents. Together, these features reflect user demand for easy and intuitive guided and visual environments for data preparation.

Figure 30 – End-user data preparation usability features
Across three years of data, attitudes toward end-user data preparation features are mostly consistent with only minor fluctuations in user priority (fig. 31). As mentioned, "immediate preview and feedback" debuted atop usability feature requirements. A different "hot button" topic, machine learning, is least relevant to respondents, although it earns respectable interest between "important" and "very important." Overall, the notion of "usability" speaks loudly about user desires and expectations.

Figure 31 – End-user data preparation usability features 2015-2017
Interest in end-user data preparation features varies somewhat by geographical regions (fig. 32). Our (small) Latin America sample leads interest in the top feature, "immediate preview and feedback," followed by North America, Asia Pacific and EMEA. Latin American respondents share top interest in other most-requested features, while EMEA respondents generally trail in interest by geography. Asia Pacific respondents report the most interest in lesser features including "visual highlighting," "automated recommendation for data relationships," "automatic transformation," and "machine learning."

![End-User Data Preparation Usability Features by Geography](image-url)

Figure 32 – End-user data preparation usability features by geography
By function, Executive Management and Sales/Marketing show the most interest in the top two usability features, "immediate preview and feedback" and "visual interface" (fig. 33). Marketing/Sales has the greatest interest in "automated detection of anomalies" and "automated data transformation." Elsewhere, functional preferences vary significantly for some features and are more clustered for others. Project/Program Management Office respondents tend to be least interested in data preparation usability features.

![End User Data Preparation Usability Features by Function](image)
Compared to other measures, interest in data preparation usability features is clustered more tightly across organizations of different sizes (fig. 34). Small organizations (1-100 employees) lead interest in several features including "immediate preview and feedback," "visual interface," automated detection anomalies," and "visual highlighting." Mid-sized organizations’ (101-1,000 employees) interest is highest in "technical expertise not required" and "support for entire data transformation process." Very large organizations (>5,000 employees) report average to slightly below-average interest in most usability features.

**Figure 34 – End-user data preparation usability features by organization size**
Interest in end-user data preparation features varies most noticeably by industry (fig. 35). Respondents in the Energy sector clearly lead interest in the top three ease-of-use features we polled (preview/feedback, visual interface, technical expertise not required). Consumer Products respondents have very high interest in automation features including "automated detection," "support for entire data transformation process," "automated recommendation," "automatically generate data transformation," and "machine learning." Healthcare respondents are most interested in "visual highlighting of relationships."

Figure 35 – End-user data preparation usability features by industry
**End-User Data Preparation Data Integration Features**

Though not quite as pronounced as usability, demand for end-user data preparation integration features is nonetheless quite strong in 2017 (fig. 36). The top three features, "access to multiple common file formats," "access to traditional databases," and "ability to combine data through joins/merges" (the most conventional integration modes) are, at minimum, "very important" to about 80 percent or more respondents. Trailing these, the ability to infer metadata is at least "important" to 78 percent of respondents. Big data and NoSQL demand are notably lower, but we can conclude user expectations for integration are undeniably high overall.

![End User Data Preparation Data Integration Features](image.png)

*Figure 36 – End-user data preparation data integration features*
Across three years of data, interest in the most conventional end-user data preparation integration features remains steady or higher, while interest in other areas declined (fig. 37). Following a 2016 dip, access to common file formats, traditional databases, and combined multiple data sets show the strongest 2017 momentum. "Ability to infer metadata" fell somewhat and, somewhat glaringly, demand for big data and NoSQL (new for 2017) integration are less relevant to users of end user data preparation tools.
Compared to other geographies, respondents in North America have slightly greater interest in "access to multiple file formats" including log files, CSV, and Excel (fig. 38). Other integration requirements vary somewhat by geography. It is interesting that sentiment for "access to big data," and especially "access to NoSQL," is strongest in Asia Pacific and Latin America, and that big data interest is lowest in North America. Asia Pacific respondents have the most interest in "ability to infer metadata."

---

**Figure 38 – End-user data preparation data integration features by geography**
Different functions/roles all share the highest interest in access to "multiple file formats" including log files, CSV, and Excel (fig. 39). BICC respondents have the highest interest in multiple integration scenarios that include "access to traditional databases," "ability to combine data across multiple data sets," and "ability to infer metadata." As we often find, Executive Management takes the most interest in potentially incipient integration opportunities including big data and NoSQL.

![End-User Data Preparation Data Integration Features by Function](image-url)

Figure 39 – End-user data preparation data integration features by function
Organizations of different sizes similarly rank end-user data preparation opportunities (fig. 40). As in other measures, the "big three" and most traditional integration scenarios universally lead interest across all organizations. Small organizations (1-100 employees) show the highest interest overall, notably in big data and NoSQL, which we might otherwise expect to be large organization opportunities.

Figure 40 – End-user data preparation data integration features by organization size
Generally speaking, different industries rank their end-user data prep integration needs similarly, though specific priorities vary from one industry to the next (fig. 41). As in other measures, the "big three" choices hold sway. In 2017, interest in flat files and ability to combine (join/merge) data is highest in Energy and Insurance. Insurance and Manufacturing lead interest in access to traditional databases, while Consumer Products organizations report above-average interest in big data.

**End-User Data Preparation Data Integration Features by Industry**

![Graph showing data integration features by industry](image)

Figure 41 – End-user data preparation data integration features by industry
End-User Data Preparation Manipulation Features
We asked organizations to gauge their interest in specific data-manipulation features and once again found a very high and broad level of interest. The top two features, "ability to aggregate and group data" and "ability to pivot data," stand out as most critical to users (fig. 42). The top seven manipulation feature priorities are all "critical" or "very important" to 60 percent up to as many as 81 percent of respondents.

Figure 42 – End-user data preparation manipulation features
Across three years of data collection, interest in end-user data manipulation features is largely constant in all areas we sampled (fig. 43). In 2017, there is a minor shift in priorities: "simple interface" and "ability to derive new data features" are slightly ahead of "ability to pivot." "Ability to aggregate data" remains the most popular manipulation. We also introduced three new manipulation choices in 2017, all of which debuted respectably but at the lower end of manipulation feature priorities.
End user data preparation manipulation features vary by geography (fig. 44). Rankings for the top six features are consistent across regions. Latin America respondents report above-average interest in features including "ability to aggregate and group" and "window/time series functions." Asia Pacific respondents have the most interest in "ability to pivot," "ability to derive new data features" and "ability to manipulate the order of data transformation steps."

**End-User Data Preparation Manipulation Features by Geography**

- Ability to aggregate and group data
- Custom user-defined functions
- Window and time series functions
- Ability to manipulate the order of data transformation steps
- Ability to pivot (convert table to matrix) and reshape (convert matrix...)
- Simple interface for imposing structure on raw data
- Ability to derive new data features from existing data (text extraction, math...)
- Ability to normalize, standardize and enrich data
- Support for cutting, merging and replacing of values

- North America
- Asia Pacific
- Europe, Middle East and Africa
- Latin America

*Figure 44 – End-user data preparation manipulation features by geography*
Interest in data manipulation features for end-user data preparation is more clustered by function (fig. 45). "Ability to aggregate and group data" is most interesting to Marketing/Sales and R&D. Involved Marketing/Sales users also prioritize "simple interface," "ability to normalize," "windows/time series," and "ability to unnest data." IT respondents are largely in the middle of interest rankings of manipulation features. Finance and Executive Management are the least interested users by function.

**Figure 45 – End-user data preparation manipulation features by function**

- Executive Management
- Information Technology (IT)
- Business Intelligence Competency Center
- Project/Program Management Office
- Marketing and Sales
- Finance
- Research and Development (R&D)
For the most part, interest in end-user data preparation manipulation features is consistent across organizations of different sizes (fig. 46). Ranking priorities are also similar with a few exceptions. Large organizations (> 1,000 employees) are more likely to seek "cutting, merging, and replacing of values." Small and mid-sized organizations report noticeably greater interest in "simple interface" and "ability to unnest data" than larger peers.

Figure 46 – End-user data preparation manipulation features by organization size
Interest in end-user data preparation manipulation feature varies by industry (fig. 47). In 2017, Energy sector respondents lead interest in all but one feature option. Consumer Products respondents have the second highest interest in "ability to pivot" but only average interest in other feature types. Insurance respondents report below-average interest in the top four feature choices. Business Services reports below-average interest in almost all features.

Figure 47 – End-user data preparation manipulation features by industry
End-User Data Preparation Supported Outputs
Respondents say the most important data prep outputs are to flat files formats, outputs to databases and direct to business intelligence tools (fig. 48). Newer, proprietary and more exotic outputs are, by comparison, unimportant to respondents. For example, users are about four times more likely to seek flat file outputs than outputs for Hadoop, a chasm that only becomes more dramatic in the case of Redshift, Azure, and other formats.

![End-User Data Preparation Supported Outputs](image)

*Figure 48 – End-user data preparation supported outputs*
The user preference for flat file, database, and BI tool format outputs for data prep extends across geographies (fig. 49). EMEA respondents have an almost universal requirement for Excel/CSV, and 90 percent or more respondents in other regions agree. A comparable if lesser common sentiment occurs across geographies for third-party business intelligence tool outputs. By comparison, interest in traditional relational database output is highest in North America and trails off noticeably in other geographies.

Figure 49 – End-user data preparation supported outputs by geography
Viewed by function, the preference for file output to Excel and CSV is virtually unanimous and overwhelming (fig. 50). However, preferences vary noticeably for other output types. Project/Program Management and Marketing/Sales have considerably greater interest in output to popular third-party BI tools. Traditional relational database output is most popular with the Project Office and Executive Management. Functional interest falls noticeably after the top three choices, where Hadoop is most interesting to Executive Management and Redshift interest is highest in the BICC.

**Figure 50 – End-user data preparation supported outputs by function**
Ninety percent or more of organizations of any size (led by mid-sized and very large organizations) share a preference for file output support of end-user data preparation (fig. 51). More than 80 percent of respondents, led by small organizations, want traditional relational database outputs. Fifty to 55 percent of organizations of different sizes want data prep outputs to BI tools, and, as we might expect, very large organizations (>5,000 employees) lead interest in outputs to Hadoop.

Figure 51 – End-user data preparation supported outputs by organization size
Vertical industries share a common preference for the "big three" end-user data preparation output types, but there are other interesting findings that vary by organization type (fig. 52). For example, Education and Business Services respondents are much more likely to want outputs to popular third-party BI tools than are respondents in Government, Healthcare, Financial Services and other verticals. Industry needs for other output formats are much lower. Though in the minority, Business Services, Financial Services and Government respondents are more likely to be Azure users.

**Figure 52 – End-user data preparation supported outputs by industry**
End-User Data Preparation Deployment Features

We asked respondents about their preferences for scheduling, monitoring, and testing aspects that make end-user data preparation more of a formal and ongoing process (fig. 53). While this resonates less so than other end user data preparation capabilities, the two most popular features, "ability to schedule execution/replay of data transformation" and "ability to iteratively sample data" are either "critical" or "very important" to more than 60 percent of respondents. Among other deployment features, interest in "API support" and "support for multiple execution environments" was less than we might have expected for data preparation deployment.

![End-User Data Preparation Deployment Features](image-url)

Figure 53 – End-user data preparation deployment features
Across three years of data, end-user data preparation deployment feature preferences rankings remain mostly constant (fig. 54). With minor shuffling, the top three deployment features remain most popular, with 2017 mean importance between "important" and "very important." In 2017, we introduced features for API support and multiple execution environments, which score respectably but not to a critical extent. That said, all features performed near or above levels of "important."

**Figure 54 – End-user data preparation deployment features 2015-2017**
Interest in end-user data preparation scheduling, monitoring, and testing features varies somewhat by geography (fig. 55). Scheduled execution of data transformation and ability to monitor ongoing data transformation are mostly uniform in importance across geographies. Push-down processing and API support have somewhat higher relevance to respondents in Asia Pacific and Latin America.

Figure 55 – End-user data preparation deployment features by geography
Sentiment toward the top three end-user data preparation deployment features is high across functions with mean interest from well above "important" to "very important" (fig. 56). Respondents in Marketing/Sales, BICC, and the Project/Program Management Office are most interested in the "ability to schedule execution/replay." Interest in "iteratively sample data" for testing transformation is highest in the Project/Program Management Office, while "push-down processing" appeals most to BICC respondents.

Figure 56 – End-user data preparation deployment features by function
Organizations of different sizes express somewhat common levels of interest in end-user data preparation deployment features, with a few noticeable differences (fig. 57). API support is more appealing to small organizations with 1-100 employees. Push-down processing is most appealing to large and small organizations, less so to mid-sized organizations with 101-1,000 employees.

Figure 57 – End-user data preparation deployment features by organization size
Interest in end-user data preparation deployment features varies by industry (fig. 58). While interest in "ability to schedule execution/replay" draws fairly steady mean scores between 3.5 and 3.9 ("important" to "very important"), variability is higher in "ability to iteratively sample data," where respondents in the Energy sector show noticeably greater interest. At the same time, push-down processing is most interesting to Insurance respondents but considerably less so to the Energy sector.

**Figure 58 – End-user data preparation deployment features by industry**
Location of End-User Data Preparation Capabilities

In 2017, we gave respondents three choices to describe their preferred deployment location scenario for end-user data preparation capabilities. Most respondents say they prefer on-premises deployment (which might include desktop, LAN, or other captive configuration inside the firewall) (fig. 59). Private cloud might include on- or off-premises single-tenant deployment. The least desirable choice is public cloud.

Figure 59 – Location of end-user data preparation capabilities
The preference for on-premises capabilities for end-user data preparation extends in near-equal sentiment across all geographies (fig. 60). In 2017, North American and EMEA respondents are least likely to support public cloud deployments. Compared to other regions, private cloud deployments are most interesting to Asia-Pacific respondents.

Figure 60 – Location of end-user data preparation capabilities by geography
Location of end-user data preparation capabilities varies more noticeably by function than by other measures (fig. 61). Perhaps with an eye on cost, Executive Management and the Project/Program Management Office are less insistent on on-premises deployment and most open to public cloud. As we would expect, Finance, for reasons of propriety, along with IT and R&D, are least likely to use public cloud. IT respondents are also most averse to private cloud deployments than other functions.

Figure 61 – Location of end-user data preparation capabilities by function
There are observable and predictable preferences for on-premises, private cloud, and public cloud deployment of end-user data prep that correlate directly to organization size (fig. 62). As organization size increases, organizations are more likely to choose on-premises deployment and less likely to pursue public cloud deployment. Perhaps most interesting is the finding that interest in private cloud is highest at smaller organizations and mostly decreases with organization size.

Figure 62 – Location of end-user data preparation capabilities by organization size
Across vertical industries, location preferences are led by on-premises and followed by private and public cloud (fig. 63). Not surprisingly, Healthcare, along with Financial Services and other regulated industries, favor on-premises deployment, though less-regulated industries including Automotive and Energy show similar intent. In 2017, only Transportation respondents have a higher sentiment for public cloud than private cloud.

Figure 63 – Location of end-user data preparation capabilities by industry
End-User Data Preparation: Standalone versus Inclusion with Other Software

We asked respondents about end-user data preparation capabilities that are included within business intelligence and/or data quality/data integration tools versus standalone data prep tools. Only 9 percent said they prefer to use end-user data preparation standalone. Sixty-five percent feel it should be part of a chosen BI tool, and 26 percent say it should be included with data quality or integration tools (fig. 64). Respondents clearly expect a seamless experience between BI and end-user data preparation (which might be a single vendor or an embedded third-party tool).

![Use of End-User Data Preparation as a Standalone Tool](image)

Figure 64 – End-user data preparation as a standalone tool
Positive user sentiment toward end-user data preparation tools/software packaged in BI tools (versus standalone) grew in 2017 (fig. 65). About two-thirds of respondents now say they prefer the BI-native inclusion of data prep tools. Sentiment for standalone tools falls below 10 percent in 2017 while interest in data prep within DQ/DI tools is flat.

Figure 65 – End-user data preparation as a standalone tool 2015-2017
Respondents in all geographies strongly prefer end-user data preparation included as a part of their business intelligence tool (fig. 66). This is especially the case in Latin America (80 percent) and Asia Pacific (71 percent). Respondents in North America and EMEA are most likely to accept data prep in DQ/DI tools, a sentiment that may extend from what's already in use. Still more than 60 percent of North American and EMEA users say BI tools are their preferred place for deploying end-user data preparation.

![Use of End-User Data Preparation as a Standalone Tool by Geography](image)

*Figure 66 – End-user data preparation as a standalone tool by geography*
Respondents across all organizational functions agree strongly that they would like end-user data preparation included as a part of their BI tool versus included with data quality/data integration or standalone (fig. 67). The least of these favorable responses comes from IT, where data prep inclusion with DQ/DI tools is highest and very likely reflects IT domain responsibilities and history of technology management.

Figure 67 – End-user data preparation as a standalone tool by function
Organizations of different sizes prefer inclusion of end-user data preparation as part of business intelligence tools (fig. 68). In 2017, this sentiment is strongest in mid-sized organizations (101-1,000 employees) and very large organizations (>5,000 employees). Small organizations with 1-100 employees and mid-sized organizations are more likely to prefer or employ standalone data prep tools, though theirs is a minority sentiment that may reflect smaller organizations’ access to more traditional/comprehensive business intelligence tools.

Figure 68 – End-user data preparation as a standalone tool by organization size
Respondents in different vertical industries universally prefer end-user data preparation included as part of business intelligence tools (fig. 69). Interest in inclusion with DQ/DI tools is highest in the Education and Transportation sectors. Standalone tools are generally least desired but most commonly found in Consumer Products, Financial Services, and Telecommunication organizations.

Figure 69 – End-user data preparation as a standalone tool by industry
Industry Support for End-User Data Preparation
Like the end-user respondent community, the provider software and services industry attaches very high importance to end-user data preparation (fig. 70). Across three years of data, industry support remains well above levels of "very important," though criticality declines somewhat in 2017. We believe this reflects maturation with end-user data preparation, increasingly a transparent component of BI tools going forward.

![Industry Importance of End-User Data Preparation 2015-2017](Image)

*Figure 70 – Industry importance of end-user data preparation 2015-2017*
Industry Support for End-User Data Preparation Usability

We asked vendors to describe their current and future support for 10 usability features associated with end-user data preparation (fig. 71). The two most supported, "immediate preview and feedback" (95 percent) and "technical expertise not required" (92 percent), are among the top three user-requested usability features (fig. 30, p. 43). Overall, industry support is good to strong across multiple capabilities. Twelve-month industry plans call for near 80 percent or greater for all usability features except machine learning.

Figure 71 – Industry support for usability features
Industry Support for End-User Data Preparation Integration

Industry investment and support for end-user data preparation integration features is robust with high levels of support for every function studied in 2017 (fig. 72). All industry participants support "access to traditional databases." There is near universal support for "access to file formats," "big data," and "ability to combine data across multiple data sets." Vendors also expect greater than 90 percent support for "infer metadata" and "NoSQL" within 12 months. Such robust support certainly answers user expectations for integration features (fig. 36, p. 49).

Figure 72 – Industry support for integration features
Industry Support for End-User Data Preparation Output Options

Industry support for output options is somewhat mixed across formats and is less robust than for integration and usability features, though mostly aligned with user demand (fig. 73). That said, industry support is mostly aligned with user preferences that favor flat files, relational databases, and popular third-party BI tools (fig. 48, p. 61). Industry support also appears to anticipate more user uptake of Hadoop, Redshift, Azure, and other outputs not critical to users today.

Figure 73 – Industry support for output options
Industry Support for End-User Data Preparation Data Manipulation Features

Industry support for data manipulation features is strong and across the board in 2017 (fig. 74). The top four features currently enjoy 90 percent or greater support, and all 11 manipulation features we sampled (with the exception of ("session-ize log/event data") will have greater than 90 percent support in future time frames. User preferences for manipulation features are somewhat aligned with industry priorities and, given high levels of current support, can be expected to meet all existing user demand (fig. 42, p. 55).

Figure 74 – Industry support for data manipulation features
Industry Support for End-User Data Preparation Deployment Features

Industry support and investment in end-user data preparation deployment features has grown to be fairly robust in 2017 (fig. 75). The most popular capability, "ability to schedule execution/replay," now stands at 85 percent, mirroring the top user priority (fig. 53, p. 66). Other user priorities are not entirely aligned. "Ability to iteratively sample data," the second most popular user feature, is currently supported by less than 70 percent of our industry sample base. Some other features strongly supported by the vendor industry (e.g., API support, push-down processing), are not top priorities for users in 2017.

Figure 75 – Industry support deployment and performance features
Industry Support for End-User Data Preparation—Cloud versus On-Premises

Industry support for end-user data preparation industry deployment options has grown across three years of study (fig. 76). In 2017, both on-premises (92 percent) and cloud (90 percent) are plainly the customers’ choice, though we are not surprised to see industry support for cloud grow at a faster pace, given the emergence of the cloud/SaaS industry. As noted earlier (fig. 59, p. 72), user demand is much stronger for on-premises deployment. Assuming users shift towards greater cloud deployment, industry support will already be in place.

Figure 76 – Industry support for cloud and on premises deployment 2015-2017
End-User Data Preparation Vendor Ratings
We include 28 vendors in our end-user data preparation ratings (fig. 77). For each vendor, we considered usability, integration, output, data manipulation, and deployment features. Only vendors that scored 50 percent or greater are included in this report.

Top-rated vendors include Trifacta (1st), Alteryx (2nd), Datawatch (tied for 3rd), Pentaho (tied for 3rd), Qlik (tied for 3rd), Datameer (tied for 4th), Microsoft (tied for 4th), Information Builders (tied for 5th), Jedox (tied for 5th), Paxata (tied for 5th), and RapidMiner (tied for 5th).

Figure 77 - End user data preparation vendor ratings
Other Dresner Advisory Services Research Reports

- Wisdom of Crowds “Flagship” Business Intelligence Market study
- Advanced and Predictive Analytics
- Analytical Data Infrastructure
- Big Data Analytics
- Business Intelligence Competency Center
- Cloud Computing and Business Intelligence
- Collective Insights®
- Enterprise Planning
- Internet of Things and Business Intelligence
- Location Intelligence
- Natural Language Analytics
- Small and Mid-Sized Enterprise Business Intelligence
- Systems Integrators
Appendix: End User Data Preparation Survey Instrument

Name*: _________________________________________________

Company Name: _________________________________________________

Address 1: _________________________________________________

Address 2: _________________________________________________

City: _________________________________________________

State: _________________________________________________

Zip: _________________________________________________

Country: _________________________________________________

Email Address*: _________________________________________________

Phone Number: _________________________________________________

Major Geography

( ) Asia/Pacific

( ) Europe, Middle East and Africa

( ) Latin America

( ) North America

What is your current title?

_________________________________________________

What function are you a part of?

( ) Business intelligence competency center

( ) Executive management
( ) Finance
( ) Information Technology (IT)
( ) Manufacturing
( ) Marketing
( ) Project/program management office
( ) Sales
( ) Research and development (R&D)
( ) Other - Write In: ________________________________

Please select an industry
( ) Advertising
( ) Aerospace
( ) Agriculture
( ) Apparel and accessories
( ) Automotive
( ) Aviation
( ) Biotechnology
( ) Broadcasting
( ) Business services
( ) Chemical
( ) Construction
( ) Consulting
( ) Consumer products
( ) Defense
( ) Distribution & logistics
( ) Education
( ) Energy
( ) Entertainment and leisure
( ) Executive search
( ) Federal government
( ) Financial services
( ) Food, beverage and tobacco
( ) Healthcare
( ) Hospitality
( ) Gaming
( ) Insurance
( ) Legal
( ) Manufacturing
( ) Mining
( ) Motion picture and video
( ) Not for profit
( ) Pharmaceuticals
( ) Publishing
( ) Real estate
( ) Retail and wholesale
( ) Sports
( ) State and local government
( ) Technology
( ) Telecommunications
( ) Transportation
( ) Utilities
( ) Other - Write In: _______________________________________________
( ) Somewhat effective
( ) Somewhat ineffective
( ) Totally ineffective

How often do users have to prepare data (e.g., combine, clean and shape datasets) to get it in a format that can be used for analysis?
( ) Constantly
( ) Frequently
( ) Occasionally
( ) Rarely
( ) Never

How often do users enrich internal data with third party data (e.g., Dun & Bradstreet, US Census)?
( ) Constantly
( ) Frequently
( ) Occasionally
( ) Rarely
( ) Never

Should end user data preparation be a standalone capability or part of another tool?
( ) Standalone
( ) Part of business intelligence tools
( ) Part of existing data quality/data integration tools
Please indicate the importance of the following usability features for end user data preparation software:

<table>
<thead>
<tr>
<th></th>
<th>Critical</th>
<th>Very important</th>
<th>Important</th>
<th>Somewhat important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical expertise/programming is &quot;NOT&quot; required to build/execute data transformation scripts</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Immediate preview and feedback for end user</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Automated recommendations for data relationships &amp; keys for combining data across multiple data sets and sources</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Visual interface for users to view and explore in-process data sets, interactively profile and refine data transformations prior to execution</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Visual highlighting of relationships between columns, attributes &amp; datasets</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Automated detection of anomalies, outliers, &amp; duplicates</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
</tbody>
</table>
Please indicate the importance of the following data integration features for end user data preparation software:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Critical</th>
<th>Very important</th>
<th>Important</th>
<th>Somewhat important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatically generate data transformation code/scripts for execution</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Support for entire data transformation process in a single application/user interface</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Machine learning and recommendations based on usage data gathered across users, groups, or organizations</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Access to traditional databases (e.g., RDBMS)</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Access to Bigdata (e.g., Hadoop)</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Access to NoSQL sources</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Access to file formats (e.g., log files, ...)</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
</tbody>
</table>
What output formats should an end user data preparation solution support?

- [ ] Traditional relational database (e.g., SQL Server)
- [ ] Excel, CSV
- [ ] Popular (third-party) business intelligence tool formats
- [ ] Hadoop
- [ ] Redshift
- [ ] Azure
- [ ] Avro
- [ ] Parquet
- [ ] Bizp/gizp
- [ ] Other - Write In: _________________________________________________

Please indicate the importance of the following data manipulation features for end user data preparation software:

<p>| Ability to infer metadata by introspecting the data elements | ( ) | ( ) | ( ) | ( ) | ( ) |
| Ability to combine data across multiple data sets and sources through joins and merging data | ( ) | ( ) | ( ) | ( ) | ( ) |</p>
<table>
<thead>
<tr>
<th>Feature</th>
<th>Critical</th>
<th>Very important</th>
<th>Important</th>
<th>Somewhat important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple interface for imposing structure on raw data</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Ability to unnest data (e.g. json/xml parsing)</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Ability to normalize, standardize &amp; enrich data</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Support for cutting, merging &amp; replacing of values</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Ability to aggregate &amp; group data</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Ability to pivot (convert table to matrix) &amp; reshape (convert matrix to table) data</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Ability to derive new data features from existing data (text extraction,</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Feature Description</td>
<td>Critical</td>
<td>Very Important</td>
<td>Important</td>
<td>Somewhat Important</td>
<td>Not Important</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------</td>
<td>----------------</td>
<td>-----------</td>
<td>--------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Ability to iteratively sample data to provide an interactive testing of transformation logic</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Push-down</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
</tbody>
</table>
### Where should end user data preparation functionality reside?

<table>
<thead>
<tr>
<th>Feature</th>
<th>Vendor 1</th>
<th>Vendor 2</th>
<th>Vendor 3</th>
<th>Vendor 4</th>
<th>Vendor 5</th>
<th>Vendor 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing of data transformations into the native data source for script execution (SQL, Pig, etc)</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>Ability to schedule the execution/replay of data transformation processing</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>Ability to monitor ongoing data transformation processing to alert on anomalies or changes in the structure</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>Support for multiple execution environments (e.g., MapReduce, Spark, Hive) based on volume and scale of data sets</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>API support (e.g., REST)</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td></td>
<td>Critical</td>
<td>Very important</td>
<td>Important</td>
<td>Somewhat important</td>
<td>Not important</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>----------------</td>
<td>-----------</td>
<td>--------------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>On-premises</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td></td>
</tr>
<tr>
<td>Private cloud</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td></td>
</tr>
<tr>
<td>Public cloud (SaaS)</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td></td>
</tr>
</tbody>
</table>