2018 Edition

End User Data Preparation

Wisdom of Crowds® Series

Licensed to Trifacta
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
In 2018, we celebrate the 11th 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.

We are also excited that our second annual conference, Real Business Intelligence, will be held June 27-28, 2018 on the MIT campus in Cambridge, Massachusetts.

Since our first market report in 2010, we added new research topics every year, and this year is no exception. In 2018, we plan to release 15 major reports, and we plan to add two new research topics: IT Analytics and Sales Planning.

This publication marks our fourth 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.

We hope you enjoy this report!

Best,

[Signature]

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**
- **Business Intelligence Competency Center**
- **Cloud Computing and Business Intelligence**
- **Collective Insights®**
- **Embedded Business Intelligence**
- **IoT Intelligence®**
- **Location Intelligence**

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 2017.

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 14th among 33 BI technologies and initiatives under our study in 2018. Perceived importance is very high over time (pp. 18-24). Industry respondent importance is high but decreasing over time (p. 83).
- Twenty-two percent say their current approach to data preparation is “highly” effective and has improved over time (pp. 25-30).
- Almost three-quarters “constantly” or “frequently” make use of end-user data preparation. Sales/Marketing and large organizations are the most frequent users (pp. 31-36).
- A large majority "constantly," "frequently," or "occasionally" enrich end-user data preparation with third-party data. Financial Services leads sector use (pp. 37-42).
- The most popular usability features include “immediate preview and feedback” and “visual interface.” Feature interest is highest in Sales/Marketing (pp. 43-48). Industry support is generally adequate for user demand (p. 84).
- The most popular integration features are access to common file formats and ability to join and merge. Big data and NoSQL are lesser priorities (pp. 49-54). Industry support is more than adequate for user demand (p. 85).
- The top two manipulation features are "ability to aggregate and group data" and "ability to pivot data." Manipulation feature interest is highest in Sales/Marketing (pp. 55-60). Industry support is well aligned with user demand (p. 87).
- The most important outputs are to flat files (Excel, CSV) and to traditional relational databases. Output options are most important to Sales/Marketing (pp. 61-65). Industry investment is well aligned with user demand (p. 86).
- The two most popular deployment features are "ability to schedule execution/replay of data transformation" and "ability to monitor ongoing data transformation" (pp. 66-71). Industry investment is in deployment and performance features is adequate but ongoing (p. 88).
- By a large margin, respondents prefer on-premises to public or private cloud deployment (pp. 72-76). Industry support for any user preference is approaching maturity (p. 89).
- While users prefer data prep capabilities within BI tools, 2018 saw a slight shift in favor of standalone tools, and capabilities as a part of existing DQ/DI tools (p. 77-82).
- End-user data preparation vendor rankings are shown on p. 90.
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-seven percent represent North America (including five Canadian provinces and the majority of U.S. states). Twenty-six percent work in EMEA; the remainder represents Asia Pacific and Latin America (fig. 1).

Geographies Represented

```
<table>
<thead>
<tr>
<th>Geographies Represented</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>67%</td>
</tr>
<tr>
<td>Europe, Middle East and Africa</td>
<td>26%</td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>5%</td>
</tr>
<tr>
<td>Latin America</td>
<td>2%</td>
</tr>
</tbody>
</table>
```

Figure 1 – Geographies represented
**Functions**

Information Technology accounts for the largest group of respondents (35 percent) by function. About 18 percent come from the Business Intelligence Competency Center (BICC). Executive Management and R&D 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.

![Functions Represented](image_url)

*Figure 2 – Functions represented*
**Vertical Industries**

Survey participants represent a wide range of vertical industries, led by Financial Services (14 percent), Technology (10 percent), and Healthcare (9 percent) (fig.3). Government and Education are the next most represented. We also 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.

![Vertical Industries Represented](image-url)

*Figure 3 – Vertical industries represented*
Organization Size
Our survey sample includes a mix of small, medium, and large organizations (fig. 4). In 2018, small organizations (1-100 employees) account for about 27 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 respondents, with very large organizations (>5,000 employees) accounting for 30 percent.

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

![Organization Sizes Represented](image)
Analysis of Findings
In 2018, (our fourth 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 2018, end-user data preparation (aka blending) ranks 14th, in the top half of 33 topics we currently study and one place up from the previous year's report (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 other 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 our fourth year of focused study of end-user data preparation, we find that 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). Seventy-two 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 4 percent say end-user data preparation is “not important.”

**Importance of End-User Data Preparation**

![Pie chart showing the importance of end-user data preparation.](image)

Figure 6 – Importance of end-user data preparation
Across four years of data, the perceived importance of end-user data preparation grew from a 2016 low to a new all-time high (fig. 7). In 2018, mean level importance stands at 3.97, or "very important." Eighty-nine percent of respondents now say end-user data preparation is, at minimum, "important." Compared to 2017, "critical" and "very important" sentiment increased while "somewhat important" sentiment decreased.

**Importance of End-User Data Preparation 2015-2018**

![Importance of End-User Data Preparation 2015-2018](image)

Figure 7 – Importance of end-user data preparation 2015-2018
By function, Sales/Marketing report the highest positive sentiment by far and the most "critical" appraisals of the importance of end-user data preparation (fig. 8). This response also describes the likely user community for the technology in 2018. After Sales/Marketing, R&D is the next most interested with combined "critical" and "very important" scores reaching 78 percent, indicating a path of future uptake of data prep technologies. That said, all functions, even Finance, report adjusted mean importance of 3.8 or higher, toward or above levels of "very important."

**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). North America reports the highest adjusted mean importance (4.0), while Asia Pacific reports slightly higher combined "critical" and "very important" scores (76 percent). EMEA respondents trail sentiment, but perceived importance is nonetheless high across all geographies.

**Figure 9 – Importance of end-user data preparation by geography**
The importance of end-user data preparation generally increases with organization size (fig. 10). In 2018, large organizations of more than 1,000 employees share the highest levels of "critical" sentiment (43 percent); very large organizations with more than 5,000 employees report the highest adjusted mean score (4.1) and the greatest combined "critical" and "very important" scores. As we found in other cases, small (generally the most agile) organizations of 1-100 employees are the next most interested, with "critical" or "very important" scores above 79 percent, ahead of mid-size peers (101-1,000 employees).

Figure 10 – Importance of end-user data preparation by organization size
Mean perceived importance of end-user data preparation in 2018 varies by industry (fig. 11). Financial Services reports the highest adjusted mean importance (4.1), just ahead of education respondents overall. However, the highest "critical" scores are reported among users in the Government (43 percent) and Technology (41 percent) industries, perhaps reflecting large user pockets of user enthusiasm (as we found in Sales/Marketing fig. 8, p. 21). Adjusted mean importance in 2018 is lowest in Manufacturing (3.5).

**Figure 11 – Importance of end-user data preparation by industry**

---

COPYRIGHT 2018 DRESNER ADVISORY SERVICES, LLC
Page | 24
Effectiveness of Current Approach to End-User Data Preparation

In 2018, the majority of organizations (52 percent) say their current end-user data preparation approach is "somewhat effective" (fig. 12). Combined with those who report "highly effective" end-user data preparation, the figure approaches nearly three-quarters (74 percent) of respondents. The remaining respondents, about 26 percent in total, report only "somewhat ineffective" or "totally ineffective" use of end-user data preparation. This positive overall response implies good levels of interaction and experience with end-user data preparation, likely in the context of increasing self-service and user autonomy.

Figure 12 – Current approach to end-user data preparation
Across four years of data collection, perceptions of end-user data preparation effectiveness steadily improve (fig. 13). The number that report "highly effective" use reaches 22 percent in 2018, while "somewhat effective" organizations account for another 52 percent. Since 2015, reports of "somewhat ineffective" or "totally ineffective" end-user data preparation declined, though improvements flattened somewhat between 2017 and 2018.

**Current Approach to End-User Data Preparation 2015-2018**

Figure 13 – Current approach to end-user data preparation 2015-2018
All functions report results in the range of "somewhat effective" mean satisfaction with end-user data preparation in 2018 (fig. 14). Respondents in the BICC and in R&D report the highest satisfaction, followed by respondents in Executive Management, IT and Sales/Marketing. These combined functions all report mean satisfaction in the adjusted mean range of 2.7 to 3.0. Satisfaction falls toward "somewhat ineffective" appraisals only in Finance.

**Current Approach to End-User Data Preparation by Function**

![Chart showing current approach to end-user data preparation by function](chart.png)

*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). Combined "highly effective" and "somewhat effective" responses are highest in North America (78 percent), though "highly effective" scores are greatest overall in Asia Pacific (31 percent). Mean scores across all geographies are closely grouped in the range of 2.85 to 2.94.

**Current Approach to End-User Data Preparation by Geography**

![Current Approach to End-User Data Preparation by Geography](figure15.png)

*Figure 15 – Current approach to end-user data preparation by geography*
Very large organizations (> 5,000 employees) and small organizations (1-100 employees) are most favorable to their current approach to end-user data preparation (fig. 16). Very large organizations have the highest adjusted mean sentiment (3.4) compared to large (1,000-5,000 employees) and mid-sized organizations (101-1,000 employees) (2.8). Despite these differences, organizations of different size all report their current approach to be in the range of "somewhat effective."

Figure 16 – Current approach to end-user data preparation by organization size
By industry, the perceived effectiveness of end-user data preparation is highest in Financial Services, Technology and Education (fig. 17). These three industries report similar grades of satisfaction, and all report "somewhat effective" data preparation practices in the adjusted mean range of 3.0. Satisfaction toward current approach drops significantly in Government and most of all in Healthcare, where fewer than 10 percent of respondents report a "highly effective" current approach.

**Current Approach to End-User Data Preparation by Industry**

![Bar chart showing the current approach to end-user data preparation by industry.](https://via.placeholder.com/150)

*Figure 17 – Current approach to end-user data preparation by industry*
Frequency of End-User Data Preparation
Seventy-two 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 unique or repeated practices, but overall usage of end-user data preparation appears to be high. Only about 7 percent of respondents only "rarely" or "never" perform end-user data preparation.

Figure 18 – Frequency of end-user data preparation
Across four years of data collection, respondents report slowly increasing frequency of use of end-user data preparation (fig. 19). In 2018, "constant" use is steady, but "frequent" use increases. "Occasional" use decreases, while "rarely" or "never" use grows slightly. At a high level, we believe there is more activity among a select and possibly growing group of users.

**Figure 19 – Frequency of end-user data preparation 2015-2018**
We would expect that issues of business performance and revenue would drive the frequency of use of end-user data preparation. In 2018, this appears to be the case with notably highest frequency of use coming from respondents in Sales/Marketing (fig. 20). (This comports with Sales/Marketing "importance" rankings (fig. 8 p. 21.) Finance is the next most active user group by function, followed in close order by R&D, the BICC, and even IT. Only Executive Management use trails off with notably fewer "constant" or "frequent" users.

Figure 20 – Frequency of end-user data preparation by function
The number of "constant" end-user data preparation users is highest in North America (30 percent) (fig. 21). Combined "constant" and "frequent" users are also highest in North America, followed by EMEA and Asia Pacific. The fewest "constant" users are in EMEA. Overall, between 88 and 94 percent of respondents in all geographic regions report, at minimum, "occasional" use of end-user data preparation.

**Frequency of End-User Data Preparation by Geography**

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

*Figure 21 – Frequency of end-user data preparation by geography*
Mean frequency of end-user data preparation in 2018 increases with organization size (fig. 22). The number of "constant" users is highest at very large organizations (>5,000 employees) followed by large organizations (1,000-5,000 employees). "Constant" or "frequent" use ranges from 69-76 percent at all organizations. Less than 10 percent of any respondent organization by size reports "rare" or "never" users.

**Figure 22 – Frequency of end-user data preparation by organization size**
End-user data preparation frequency in 2018 is highest at Financial Services organizations, where nearly half of respondents report "constant" use (fig. 23). All other industries report fewer than half as many "constant" users as found in Financial Services. Government respondents are the next most frequent users, though only 10 percent or fewer respondents in all other industries only "rarely" perform end-user data preparation. The lowest mean score of any industry, found in technology (3.74), still reflects "frequent" use of end-user data preparation.

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

A large majority of organizations "constantly," "frequently," or "occasionally" enrich end-user data preparation with third-party data (fig. 24). Only about one-third of users "rarely" or "never" use external data. Still, just 6 percent are "constant" users of non-proprietary data, perhaps indicating that requirements for third-party data are often circumstantial or tied to unique users, projects, or queries.

**Figure 24 – Frequency of end-user data preparation enrichment with third-party data**
Across four years of data, respondents have reported shifting patterns but consistent overall frequency of the use of third-party data use in conjunction with end-user data preparation (fig. 25). While reports of "constant" use most notably decline during our study, adjusted mean use in 2015 (2.94) is identical to what is reported in 2018. Also, minimally "occasional" use of external data is virtually unchanged over time, while "rarely" and "never" use declines. At a mean slightly below "occasional" use, we continue to observe that organizations most often grapple with internal data.

**Figure 25 – Frequency of end-user data preparation enrichment with third-party data 2015-2018**
By function, R&D maintains a slight edge over Executive Management, the BICC, IT, and Sales/Marketing in frequency of enriching end-user data preparation with third-party data (fig. 26). Outside of Finance, where frequency dips sharply, we observe that third-party data enrichment is an "occasional" cross-functional pursuit and not a front-burner priority in 2018.

![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 greatest in Asia Pacific, followed by North America and EMEA (fig. 27). Respondents in Asia Pacific are 76 percent likely to be, at minimum, "occasional" users of third-party data enrichment. However, the drop-off in occasional third-party data enrichment in other regions is not extreme. Mean sentiment toward third-party data use is again generally near the 3.0 or "occasional" level of use globally.

**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 not consistent or linear across organizations of different sizes (fig. 28). Very large organizations (>5,000 employees) and mid-sized organizations (101-1,000 employees) are the most frequent users of third-party data and show similar "constant," "frequent," and "occasional" user profiles. Small organizations (1-100 employees) are the next most likely to enrich with third-party data. Among those large organizations with 1,001 to 5,000 employees, 45 percent "never" or "rarely" enrich the process with third-party data.

**Figure 28 – Frequency of end-user data preparation enrichment with third-party data by organization size**

COPYRIGHT 2018 DRESNER ADVISORY SERVICES, LLC

Page | 41
By industry, highly transactional and fraud-sensitive Financial Services/Insurance respondents are most likely to be frequent users of third-party data enrichment (fig. 29). More than 80 percent in this group are at least "occasional" users. Technology industry respondents are the next most likely to enrich with third-party data. Use thereafter declines across the Education, Healthcare, and Government verticals.

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

2018 respondents have strong interest in a wide range of end-user data preparation usability features, all of which they consider "important" or "very important" (fig. 30). We believe this reflects good understanding of needs and high expectations for data preparation features, led by utilitarian features. "Immediate preview and feedback" and "visual interface" lead requirements with upwards of 70 percent describing these features as "very important" at minimum. Every feature, with the exception of (much-hyped) machine learning, is considered at least "important" to 90 percent or more of respondents.

End-User Data Preparation Usability Features

Machine learning and recommendations based on usage data gathered across users, groups, or organizations
Automated recommendations for data relationships & keys for combining data across multiple data sets and...
Support for entire data transformation process in a single application/user interface
Automatically generate data transformation code/scripts for execution
Visual highlighting of relationships between columns, attributes & datasets
Technical expertise/programming is *NOT* required to build/execute data transformation scripts
Automated detection of anomalies, outliers, & duplicates
Visual interface for users to view and explore in-process data sets, interactively profile and refine data...
Immediate preview and feedback for end user

Figure 30 – End-user data preparation usability features
Across four years of study, attitudes toward end-user data preparation features are mostly unchanged by rank (fig. 31). As mentioned, "immediate preview and feedback" (added in 2017), and "visual interface" retain top ranks year over year. Interest in all features, with the exception of "machine learning," declines somewhat compared to 2017, most notably "technical expertise not required, "automated recommendations" and "support for entire data transformation process."

**End-User Data Preparation Usability Features 2015-2018**

- Immediate preview and feedback for end user
- Visual interface for users to view and explore in-process data sets, interactively profile and...
- Automated detection of anomalies, outliers, & duplicates
- Visual highlighting of relationships between columns, attributes & datasets
- Technical expertise/programming is *NOT* required to build/execute data transformation...
- Automated recommendations for data relationships & keys for combining data across...
- Support for entire data transformation process in a single application/user interface
- Automatically generate data transformation code/scripts for execution
- Machine learning and recommendations based on usage data gathered across users, groups,...

![Figure 31 – End-user data preparation usability features 2015-2018](image-url)
By function, Sales/Marketing shows the most interest in the top three usability features, "immediate preview and feedback," "visual interface," and "automated detection of anomalies" (fig. 32). Sales/Marketing respondents also are interested in automation and simplicity features (e.g., "automatically generate data transformation" and "support for entire data transformation process in a single app"). Finance and Executive Management respondents lead interest in "technical expertise not required" solutions to lessen help requirements.

The diagram illustrates the end-user data preparation usability features by function, highlighting the top three interests: "immediate preview and feedback," "visual interface," and "automated detection of anomalies." The diagram also shows other features of interest, such as automation and simplicity features like "automatically generate data transformation code/scripts for execution."
Interest in end-user data preparation features varies somewhat by geographical regions (fig. 33). Interest in several features including "immediate preview and feedback" and "technical expertise not required" is well clustered among different geographies. Asia-Pacific respondents report the keenest interest in most features including "visual interface," "automated detection of anomalies," as well as the greatest interest in "machine learning." EMEA respondents have the lowest interest in all usability features sampled.

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

![Diagram showing the interest levels of various usability features by geography.](image)

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

*Figure 33 – End-user data preparation usability features by geography*
Compared to other measures, interest in data preparation usability features is consistently clustered across organizations of different sizes (fig. 34). Interest in usability features appears to slightly increase with organization headcount as well. Very large organizations (>5,000 employees) narrowly lead "immediate preview and feedback," "visual interface," and "technical expertise not required." Large organizations with 1,001 to 5,000 employees similarly edge out others in "automated detection of anomalies" and "visual highlighting, along with all remaining automation features and "machine learning." Small organizations (1-100 employees) generally trail interest in usability features, but by a narrow margin.

**End-User Data Preparation Usability Features by Organization Size**

- **Immediate preview and feedback for end user**
- **Visual interface for users to view and explore in-process data sets, interactively profile and refine data transformations prior to…**
- **Automated detection of anomalies, outliers, & duplicates**
- **Technical expertise/programming is *NOT* required to build/execute data transformation scripts**
- **Automatically generate data transformation code/scripts for execution**
- **Support for entire data transformation process in a single application/user interface**
- **Automated recommendations for data relationships & keys for combining data across multiple data sets and…**
- **Machine learning and recommendations based on usage data gathered across users, groups, or organizations**
- **Visual highlighting of relationships between columns, attributes & datasets**

![Figure 34 – End-user data preparation usability features by organization size](image-url)
Compared to other measures, interest in end-user data preparation usability features varies most noticeably by industry (fig. 35). Respondents in Financial Services have the greatest interest in "immediate preview and feedback" and take strong interest in "visual highlighting." In 2018, Education and Technology respondents tend to place above-average importance on multiple features, while Healthcare and Manufacturing respondents mostly report below-average usability feature interest.

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

- Immediate preview and feedback for end user
- Visual interface for users to view and explore in-process data sets, interactively profile and refine data transformations prior to...
- Automated detection of anomalies, outliers, & duplicates
- Technical expertise/programming is *NOT* required to build/execute data transformation scripts
- Visual highlighting of relationships between columns, attributes & datasets
- Support for entire data transformation process in a single application/user interface
- Automatically generate data transformation code/scripts for execution
- Automated recommendations for data relationships & keys for combining data across multiple data sets and sources
- Machine learning and recommendations based on usage data gathered across users, groups, or organizations

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

Though not as strongly pronounced as usability, demand for end-user data preparation integration features is nonetheless strong in 2018 (fig. 36). The top two features, "access to file formats" and "ability to combine data across multiple data sets and sources" are seen as "critical" to more than half of respondents. "Access to traditional databases" follows closely and is either "critical" or "very important" to more than three-quarters of respondents. "Ability to infer metadata" is "critical" or "very important" to more than half of respondents. Big data and NoSQL demand are notably lower, but still, at minimum, "important" to large majorities of users.

Figure 36 – End-user data preparation data integration features
Year-over-year interest in end-user data preparation integration features is for the most part steady or slightly higher in 2018 (fig. 37). The leading features, "access to file formats" and "ability to combine multiple data sets," are mostly unchanged, "access to traditional databases" slips slightly. Remaining categories including big data and NoSQL increase slightly. Across all four years of our study, rankings remain mostly consistent around a dip in 2016.

**End-User Data Preparation Data Integration Features 2015-2018**

![Graph showing data integration features from 2015 to 2018](image)

*Figure 37 – End-user data preparation data integration features 2015-2018*
Different functions/roles are tightly clustered around demand for the top integration feature, "access to multiple file formats" (e.g., log files, CSV, and Excel) (fig. 38). Sales and marketing have the greatest interest in "ability to combine data across multiple data sets." BICC respondents are most likely to require multiple integration scenarios that include "access to traditional databases." R&D has high interest in big data and NoSQL, though Sales/Marketing interestingly shares above-average interest in these categories.

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

![End-user data preparation data integration features by function](image)

*Figure 38 – End-user data preparation data integration features by function*
With a slight edge to North-American respondents, all geographies share similarly high interest in the top three integration end-user data preparation integration features (fig. 39). Elsewhere, Asia Pacific respondents most notably show standout interest in "access to NoSQL sources" and lead interest in "ability to infer metadata," while EMEA respondents report above-average interest in big data.

End-User Data Preparation Data Integration Features by Geography

Figure 39 – End-user data preparation data integration features by geography
Interest in end-user data preparation integration features rises incrementally with organization size (fig. 40). As in other measures, the "big three" and most traditional integration scenarios universally lead interest across all organizations. By a larger margin, interest in big data is highest among the largest organizations (> 5,000 employees), while NoSQL interest is greatest at mid-sized organizations (101-1,000 employees) followed by small (1-100 employees) and very large organizations.

**End-User Data Preparation Data Integration Features by Organization Size**

![Bar chart showing end-user data preparation data integration features by organization size.](image)

*Figure 40 – End-user data preparation data integration features by organization size*
Interest in the top three end-user data preparation integration features is also highest across industries we sampled for 2018 (fig. 41). (Among these three features, interest hovers above or near an adjusted mean of 4.0 or "very important.") Interest in "ability to infer metadata" is also somewhat clustered. Financial Services and Technology respondents assign the highest importance to big data and NoSQL integration.

---

**Figure 41 – End-user data preparation data integration features by industry**
End-User Data Preparation Manipulation Features

We asked organizations to score their interest in specific data-manipulation features and once again found a very high and broad level of interest (fig. 42). All but the lowest-ranked feature (session-ize log or event data) was at least "very important" to 50 percent or far more respondents. The top two features, "ability to aggregate and group data" and "ability to pivot data," are most critical and reflective of classic spreadsheet operations. The next most important features, "simple interface," "ability to derive new data features," and "ability to normalize, standardize, and enrich data," are likewise attuned to finding improvements to common data manipulations.

Figure 42 – End-user data preparation manipulation features
Year-over-year interest in end-user data preparation manipulation features is mostly steady (fig. 43). Priority rankings over the last 12 months are effectively unchanged as well. There are slight upticks in demand for "ability to standardize, normalize, and enrich" and "ability to un-nest data." Interest in "window and time-series functions" decreases slightly. Across our four years of data collection, sentiment is mostly unchanged around a dip in 2016.

**End-User Data Preparation Manipulation Features 2015-2018**

- Session-ize log or event data
- Ability to unnest data (e.g. json/xml parsing)
- Custom user defined functions
- Window and time series functions
- Ability to manipulate the order of data transformation steps
- Support for cutting, merging & replacing of values
- Ability to normalize, standardize & enrich data
- Ability to derive new data features from existing data (text extraction, math…)
- Simple interface for imposing structure on raw data
- Ability to pivot (convert table to matrix) & reshape (convert matrix to table) data
- Ability to aggregate & group data

*Figure 43 – End-user data preparation manipulation features 2015-2018*
Interest in data manipulation features for end-user data preparation varies rather broadly by function (fig. 44). Sales/Marketing stands out as being most interested in nine of 11 functions we sampled, including five of the top six. Finance expresses the most interest in "simple interface" but has below-average interest in most other features. BICC respondents have the most interest in "ability to manipulate the order of data transformation steps" and rank it among the top two or three functions for all features we sampled.

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

Figure 44 – End-user data preparation manipulation features by function
End-user data preparation manipulation features vary by geography (fig. 45). Interest in six of seven top features, including "ability to aggregate," "ability to pivot," "ability to normalize," and "ability to derive," is highest among North American respondents. Asia-Pacific respondents lead interest in "simple interface" and lesser features including "custom user-defined functions" and "window and time series functions." EMEA respondents report the least interest in manipulation features by geography.
Though results are somewhat clustered, very large organizations (> 5,000 employees) lead overall interest in end-user data preparation manipulation features (fig. 46). Generally, interest in these features declines with organization size, and small organizations (1-100 employees) are the least interested overall. Feature rankings by order are also mostly steady across different-sized organizations.

**End-User Data Preparation Manipulation Features by Organization Size**

![End-User Data Preparation Manipulation Features by Organization Size](image)

**Figure 46** – End-user data preparation manipulation features by organization size
Interest in end-user data preparation manipulation features varies by industry, though some feature scores are tightly clustered (fig. 47). In 2018, Financial Services and Healthcare respondents score narrowly higher for the top three features ("ability to aggregate," "ability to pivot," and "simple interface"). Government respondents show the greatest interest in "ability to derive new data" while Technology respondents have high interest in "ability to normalize" and the most interest in "ability to un-nest."

End-User Data Preparation Manipulation Features by Industry

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 file formats (Excel, CSV) and to traditional relational databases (sentiments that also hold over time) (fig. 48). Some other common outputs and many newer ones are, by comparison, relatively unimportant in 2018. For example, users are about three times more likely to seek flat file outputs than outputs for Hadoop. The difference is more than four times when Excel is compared to Spark, a chasm that only becomes more dramatic in the case of Redshift, Azure, and other formats.

Figure 48 – End-user data preparation supported outputs
Viewed by function, Sales/Marketing reports the most interest in nearly all end-user data preparation supported outputs (fig. 49). After Sales, BICC, IT, and Finance are also highly likely to require the most supported Excel/CSV output types. Apart from Excel, Finance is generally uninterested in multiple supported outputs. R&D interest picks up in areas of traditional databases (JSON, Hadoop, and Spark) but is lower elsewhere. Executive Management is broadly disinterested in support for end-user data preparation outputs.

**End-User Data Preparation Supported Outputs by Function**

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

*Figure 49 – End-user data preparation supported outputs by function*
The user preference for flat file Excel and CSV outputs for data prep is greatest in Asia Pacific but extends universally across geographies (fig. 50). Traditional relational database outputs are a strong second choice across all geographies. Thereafter, North American and EMEA respondents are far more likely than Asia-Pacific counterparts to require popular third-party BI tool outputs. JSON output is more equally popular across geographies. Among trailing choices, Hadoop output is most popular with EMEA respondents, while Access and Azure preference is strongest in Asia Pacific.

End-User Data Preparation Supported Outputs by Geography

Figure 50 – End-user data preparation supported outputs by geography
Ninety percent or more of organizations of any size (led by large and very large organizations) share the highest preference for flat file output support of end-user data preparation (fig. 51). Again, traditional relational database output support is the strong second choice, led by respondents in large organizations (1,001-5,000 employees). With the exception of JSON, where mid-sized organizations (101-1,000 employees) show the most interest, large and very large organizations lead demand for all types of data prep output support.

**End-User Data Preparation Supported Outputs by Organization Size**

![End-user data preparation supported outputs by organization size](image-url)

*Figure 51 – End-user data preparation supported outputs by organization size*
All industries we sampled in 2018 share the greatest preference for Excel/CSV outputs and the second choice of traditional relational database output support (fig. 52). Government respondents show the greatest interest in Excel/CSV but also the least interest among industries in relational database output support. Healthcare respondents especially show affinity for relational database output support compared to all other formats. Technology organizations are most attracted to JSON output support. Financial Services is most interested in Hadoop, while Manufacturing respondents are most likely to use Access.

**End-User Data Preparation Supported Outputs by Industry**

![End-user data preparation supported outputs by industry](image)

*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 part of a more formal ongoing process (fig. 53). While such features resonate less with respondents compared to other end-user data preparation capabilities, the two most popular features ("ability to schedule execution/replay of data transformation" and "ability to monitor ongoing data transformation processing") are either "critical" or "very important" to 60 percent or more of respondents. Among other deployment features, interest in "API support" gathers above-average "critical" interest in certain organizations, while "support for multiple execution environments" such as MapReduce, Spark, or Hive is less important to organizations in 2018.

End-User Data Preparation Deployment Features

![Diagram showing the preference levels for various deployment features such as support for multiple execution environments, push-down processing, API support, iterative sampling, and ability to schedule execution/replay.]

Figure 53 – End-user data preparation deployment features
Between 2017 and 2018, interest in end-user data preparation deployment features is largely steady with only minor changes, led by upticks in API support and "support for multiple execution environments" (fig. 54). Across our four years of data, rankings remain mostly steady, led consistently by "ability to schedule," "ability to monitor," and "ability to iteratively sample."

### End-User Data Preparation Deployment Features 2015-2018

<table>
<thead>
<tr>
<th>Feature</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to schedule the execution/replay of data transformation processing</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Ability to monitor ongoing data transformation processing to alert on anomalies or changes in the structure</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>Ability to iteratively sample data to provide an interactive testing of transformation logic</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>API support (e.g., REST)</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Push-down processing of data transformations into the native data source for script execution (SQL, Pig, etc.)</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Support for multiple execution environments (e.g., MapReduce, Spark, Hive) based on volume and scale of data sets</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Figure 54 – End-user data preparation deployment features 2015-2018
Sentiment toward the top three end-user data preparation deployment features is high across functions with mean interest ranging from above "important" to "very important" (fig. 55). Interest in all deployment features is consistently highest among Sales/Marketing respondents. Finance most highly ranks "ability to schedule." IT is the second most likely function to require "ability to monitor ongoing data transformation." BICC and R&D respondents report above-mean interest in all deployment features, indicating a runway for ongoing and future use.

**End-User Data Preparation Deployment Features by Function**

- **Ability to schedule the execution/replay of data transformation processing**
- **Support for multiple execution environments (e.g., MapReduce, Spark, Hive) based on volume and scale of data sets**
- **Push-down processing of data transformations into the native data source for script execution (SQL, Pig, etc.)**
- **API support (e.g., REST)**
- **Ability to monitor ongoing data transformation processing to alert on anomalies or changes in the structure**
- **Ability to iteratively sample data to provide an interactive testing of transformation logic**

*Figure 55 – End-user data preparation deployment features by function*
Interest in end-user data preparation scheduling, monitoring, and testing features is only somewhat consistent by geography (fig. 56). Most notably, Asia-Pacific respondents report well above-mean interest in API support and above-mean interest in "support for multiple execution environments." Asia-Pacific respondents’ interest in "ability to monitor ongoing data transformation" is slightly below that of other regions, as is EMEA respondents’ interest in "ability to iteratively sample data."

End-User Data Preparation Deployment Features by Geography

Figure 56 – End-user data preparation deployment features by geography
Interest in end-user data preparation deployment features generally increases in linear fashion with organization headcount (fig. 57). In 2018, the exception to this rule is API support, which is most popular in small organizations (1-100 employees) and becomes less important as organization size grows.

**Figure 57 – End-user data preparation deployment features by organization size**
Interest in end-user data preparation deployment features varies most erratically by industry (fig. 58). In 2018, Government respondents report the highest interest in "ability to schedule" but have below-average interest in "push-down processing" and "support for multiple execution environments." Elsewhere, technology respondents report above-average interest in all deployment features, most notably "API support." Financial Services leads interest in "ability to iteratively sample" and shares (with Technology respondents) the highest interest in "push-down processing" and "support for multiple execution environments.

**Figure 58 – End-user data preparation deployment features by industry**
Location of End-User Data Preparation Capabilities
We gave respondents three choices to describe their preferred deployment location scenario for end-user data preparation capabilities in 2018 (fig. 59). By a large margin, respondents prefer on-premises deployment (which might include desktop, LAN, or other configurations inside the firewall). Compared to on-premises deployments, which are "critical" or "very important" to more than two-thirds of respondents, private cloud deployments are "critical" or "very important" to fewer than half of our sample. Public cloud sentiment is the least-regarded option, considered "not important" or only "somewhat important" to more than half of the sample.

Figure 59 – Location of end-user data preparation capabilities
All functions we sampled in 2018 prefer location of end-user data preparation capabilities to be on premises (fig. 60). IT and R&D respondents are most adamant in this regard, and BICC respondents are the next most likely to prefer on-premises deployment. Responses are most tightly clustered among Executive Management respondents, who are also most likely to accept public cloud deployment, perhaps with an eye toward lower cost.

**Location of End-User Data Preparation Capabilities by Function**

![Image of spider diagram showing the location of end-user data preparation capabilities by function.]

Figure 60 – Location of end-user data preparation capabilities by function
The preference for on-premises capabilities for end-user data preparation extends in near-equal sentiment across all geographies (fig. 61). In 2018, Asia-Pacific respondents have by far the most interest in private cloud (equal to Asia-Pacific on-premises sentiment) and the highest interest in public cloud, for which sentiment declines across North American and EMEA respondents.

**Location of End-User Data Preparation Capabilities by Geography**

*Figure 61 – Location of end-user data preparation capabilities by geography*
There are observable and predictable preferences for on-premises, private cloud, and public cloud deployment of end-user data prep that correlate to organization size (fig. 62). Sentiment is always highest for on-premises deployment; but as organization size increases, organizations are more likely to choose on-premises deployment and less likely to pursue public cloud deployment. Also, sentiment toward private cloud tends to decrease (if not as consistently) as organization size increases.

**Location of End-User Data Preparation Capabilities by Organization Size**

*Figure 62 – Location of end-user data preparation capabilities by organization size*
Across vertical industries, on-premises end-user data preparation location preferences lead in every case, most dramatically in Government, Healthcare, and Financial Services organizations (fig. 63). However, Technology respondents are almost equally likely to report private cloud as on-premises deployments. Technology organizations are also most likely to report public cloud end-user data preparation deployments, though this is consistently the last choice among organizations in any industry.

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

We asked respondents to describe their preference for end-user data preparation capabilities that are included within business intelligence and/or data quality/data integration tools versus stand-alone data prep tools (fig. 64). They report their greatest preference is data prep as a part of BI tools, followed by data prep in DQ/DI tools (28 percent). Only 17 percent say they prefer using end-user data preparation stand-alone tools. These findings might also reflect respondents' preference for 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 Stand-alone Tool

![Pie chart showing preferences for end-user data preparation](image-url)
Positive user sentiment toward end-user data preparation tools/software packaged in BI tools (versus stand-alone) is the leading choice across our four years of study (fig. 65). In 2018, however, sentiment toward stand-alone tools almost doubled from 9 to 17 percent, largely at the expense of data prep as a part of BI tools and platforms. Over our four years of data, the preference for data preparation within DQ and DI tools is steady at 26 to 28 percent.

Use of End-User Data Preparation as a Stand-alone Tool 2015-2018

Figure 65 – End-user data preparation as a stand-alone tool 2015-2018
Respondents across nearly all organizational functions agree they would like end-user data preparation included as a part of their BI tool versus within DI/DQ or stand-alone (fig. 67). This sentiment is strongest in Finance and the BICC. The curious exception to this preference is R&D, where data preparation rises more often within data integration and data quality tools. Stand-alone tools are the third choice across all functions sampled.

**Use of End-User Data Preparation as a Stand-alone Tool by Function**

![Diagram showing the use of end-user data preparation as a stand-alone tool by function.](image)

*Figure 66 – End-user data preparation as a stand-alone tool by function*
Majorities of respondents in all geographies prefer end-user data preparation included as a part of their business intelligence tool (fig. 68). This is most often the case in EMEA and Asia Pacific. Asia-Pacific respondents also have the highest interest in stand-alone tools and, contrary to trend, prefer stand-alone tools to data preparation within existing data quality or data integration tools.

Use of End-User Data Preparation as a Stand-alone Tool by Geography

![Bar chart showing the use of end-user data preparation as a stand-alone tool by geography. The chart indicates the percentage of respondents in Asia Pacific, North America, and Europe, Middle East, and Africa who prefer data preparation as part of business intelligence tools, part of existing data quality/data integration tools, and as stand-alone tools.](chart.png)

Figure 67 – End-user data preparation as a stand-alone tool by geography
Organizations of different sizes most prefer inclusion of end-user data preparation as part of business intelligence tools (fig. 68). In 2018, this sentiment is strongest in mid-sized organizations (101-1,000 employees) and very large organizations (>5,000 employees). Large organizations of 1,001 to 5,000 employees are most likely to employ data preparation within existing data quality or data integration tools. Small organizations with 1-100 employees are somewhat more likely than others to use stand-alone data preparation.

**Figure 68 – End-user data preparation as a stand-alone tool by organization size**
Respondents in different vertical industries most prefer end-user data preparation included as part of business intelligence tools (fig. 69). This is most strongly the case among Education respondents. However, Government respondents are equally likely to use stand-alone or DQ/DI-inclusive data prep as those that are a part of BI tools. Technology and Healthcare organizations are most likely to use stand-alone data preparation tools.

**Use of End-User Data Preparation as a Stand-alone Tool by Industry**

![Figure 69 – End-user data preparation as a stand-alone tool by industry](image-url)
Industry Support for End-User Data Preparation

Like the end-user respondent community, the provider software and services industry continues to attach high importance to end-user data preparation (fig. 70). That said, it appears that parts of the industry may feel data prep “crossed the chasm” since criticality and mean levels of importance decline over the last two years. This year, reports of “not important” also ticked up from 3 to 8 percent. In 2018, adjusted mean level importance stands at 3.4 (compared to an all-time high of 3.8), a drift toward levels of “important” versus "very important" overall. While other BI imperatives may be coming more to the fore, we are confident that end-user data preparation will be a common and transparent component or feature of BI tools going forward.

Industry Importance of End-User Data Preparation 2015 to 2018

Figure 70 – Industry importance of end-user data preparation 2015-2018
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). Generally, we can report that the industry does a good job at providing adequate support for user demand. The most-supported feature today, "immediate preview and feedback" (90 percent), is also the top user priority (fig. 31, p. 44). The next two top user priorities ("visual interface" and "automated detection of anomalies") are less supported today, though vendors expect to address shortfalls in the next 12 months. Like users, industry respondents are still in the process of addressing machine learning capabilities, though vendors are driving the technology forward in 2018 and beyond.

![Industry Support for Usability Features](image)

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 2018 (fig. 72). Nearly all industry participants already 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).

![Industry Support for Integration Features](#)

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

Industry support for output options is less robust than for integration and usability features but appears to meet the “80-20 rule” in addressing market demand (fig. 73). The five most-supported outputs match the top five user requirements (fig. 48, p. 61) in slightly different order. Industry support also appears to anticipate more user uptake of Hadoop, Redshift, Azure, and other outputs that are not broadly demanded by 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 2018 (fig. 74). The top six features currently enjoy 80 percent or greater support, and all 11 manipulation features we sampled will surpass 80 percent support within 12 months. The top seven industry-supported manipulation features also match the top seven user priorities (fig. 42, p. 55) in only slightly different order.

**Industry Support for Data Manipulation Features**

![Bar chart showing industry support for data manipulation features](image)

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 is robust, and additional investment will extend capabilities in 2018 (fig. 75). Top user priorities (fig. 53, p. 66), including “ability to schedule” and “ability to iteratively sample,” are among the top three industry-supported features today. The second-to-top user requirement, “ability to monitor,” is least supported by vendors today, but vendors expect to address that gap with future investment.

**Industry Support for Deployment and Performance Features**

- Ability to monitor ongoing data transformation processing to alert on anomalies or changes in the structure
- Support for multiple execution environments (e.g., MapReduce, Spark, Hive) based on volume and scale of data sets
- Push-down processing of data transformations into the native data source for script execution (SQL, Pig, etc.)
- Ability to iteratively sample data to provide an interactive testing of transformation logic
- Ability to schedule the execution/replay of data transformation processing
- API support (e.g., REST)

![Industry Support Chart]

*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 is approaching maturity in 2018 (fig. 76). Currently, nearly 90 percent of vendors support or plan to support on-premises deployment; about 85 percent currently support cloud deployment. Vendors expect future investment to make cloud deployment universally available within 24 months. As noted earlier (fig. 59, p. 72), user demand is much stronger for on-premises versus public or private cloud deployment. Assuming users shift towards greater cloud deployment, industry support will already be in place.

![Industry Support for Cloud and On-premises Deployment](image)

*Figure 76 – Industry support for cloud and on-premises deployment*
**End-User Data Preparation Vendor Ratings**

We include 28 vendors in our end-user data preparation ratings (fig. 77). For each vendor, we consider usability, integration, output, data manipulation, and deployment features. Only vendors that score 50 percent or greater are included in this report.

Top-rated vendors include Trifacta in 1st place, Datameer in 2nd, ClearStory Data and Datawatch tied for 3rd, Hitachi Vantara (Pentaho) in 4th and RapidMiner in 5th place.

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

- “Flagship” Wisdom of Crowds® Analytical Data Infrastructure Market Study
- “Flagship” Wisdom of Crowds® Business Intelligence Market Study
- “Flagship” Wisdom of Crowds® Enterprise Planning market Study
- Advanced and Predictive Analytics
- Big Data Analytics
- Business Intelligence Competency Center
- Cloud Computing and Business Intelligence
- Collective Insights®
- Data Catalog
- Embedded Business Intelligence
- IoT Intelligence®
- Location Intelligence
- Small and Mid-Sized Enterprise BI
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: ________________________________________________

How many employees does your company employ worldwide?

( ) 1 - 100

( ) 101 - 1000

( ) 1001 - 5000

( ) More than 5000

How important is it for users to be able to prepare data (e.g., combine, clean, shape datasets) prior to analysis?*

( ) Critical

( ) Very important

( ) Important

( ) Somewhat important

( ) Not important

What tool(s) do users currently use to prepare data for analysis?

____________________________________________

____________________________________________

____________________________________________

____________________________________________

How effective is the current approach to end user data preparation for Business Intelligence/user analysis today?

( ) Highly effective
( ) 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 <em>NOT</em> 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>
<tr>
<td>Feature</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>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>
</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>Access to traditional databases (e.g., RDBMS)</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>Access to big data (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</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:

<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 un-nest 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</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
</tbody>
</table>
Please indicate the importance of the following deployment 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>Ability to iteratively sample data to provide an interactive testing of transformation</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>logic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Push-down 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>
</tr>
<tr>
<td>Ability to schedule the execution/replay of data transformation processing</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>
</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>
</tr>
<tr>
<td>API support (e.g., REST)</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></th>
<th>Critical</th>
<th>Very important</th>
<th>Important</th>
<th>Somewhat important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-premises</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>
</tr>
<tr>
<td>Public cloud (SaaS)</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
</tbody>
</table>