



We're the 2 in B2B<sup>SM</sup>

***The Download Reporter***<sup>TM</sup>

**Data Dictionary and Methodology**

Universal Data Interface Corporation Technical Document

VERSION 1.0a, January 30, 2003

## Copyright Notice and Disclaimer

© Copyright Universal Data Interface Corporation ("UDICo") 2003. All rights reserved. Material that is quoted from sources other than UDICo is the property of the copyright holder. Company and product names are trademarks or registered trademarks of their respective companies.

### **No Affiliation to CNet or Download.com**

Universal Data Interface Corporation and The Download Reporter™ have no affiliation with CNet or Download.com. The information contained in this document is based upon the observation of information posted on Internet web sites.

### **Data is Not Audited and is Provided "As Is"**

While every effort has been made to ensure the accuracy of this information, this document and the reports distributed by The Download Reporter™ may contain errors. UDICo does not warrant the accuracy of this information, nor has this information been audited, certified or endorsed by any third party. This information is provided "as-is" and any use, application or interpretation of this data is at the sole risk and liability of the customer and/or licensor.

### **Additional Terms and Conditions**

The Download Reporter™ End User License Agreement ("EULA") may be accessed from the UDICo web site at [www.udico.com/reports/eula.html](http://www.udico.com/reports/eula.html). Use of this documentation and the information provided by The Download Reporter™ is governed by the terms and conditions contained in the EULA.

## Table of Contents

<b>1. PRODUCT MISSION AND METHODOLOGY .....</b>	<b>1</b>
1.1. Why The Download Reporter™? .....	1
1.2. How to Use this Data .....	1
1.3. Data Gathering Methodology.....	1
1.4. Calculating Economic Value.....	2
<b>2. QUANTITATIVE MARKETING .....</b>	<b>4</b>
2.1. How to Use the Download Reporter™ .....	4
2.2. Have Realistic Expectations .....	4
2.3. Listen to the Numbers .....	4
2.4. Try the Products and Form Your Own Opinion.....	4
2.5. Do Your Own Content Analysis .....	4
2.6. Look for Correlated Success Factors .....	5
2.7. Putting it All Together.....	5
<b>3. OVERVIEW OF DATA STRUCTURES.....</b>	<b>6</b>
3.1. Reporting Tables Summary .....	6
3.2. Notes on Selected Fields.....	6
3.3. Title Rank, Quartile and Percentiles .....	8
3.4. Ranks by Category and Subcategory.....	8
<b>4. THE STAT TABLE DATA DICTIONARY .....</b>	<b>9</b>
<b>5. THE STATVALUE TABLE DATA DICTIONARY.....</b>	<b>10</b>
<b>6. THE CATEGORYREPORT TABLE .....</b>	<b>11</b>
<b>7. THE CATEGORYVALUE TABLE .....</b>	<b>12</b>

<b>8. UDICO CONTACT INFORMATION</b> .....	<b>13</b>
<b>8.1. Company Overview</b> .....	<b>13</b>
<b>8.2. Contact Information</b> .....	<b>13</b>

# 1. Product Mission and Methodology

## 1.1. *Why The Download Reporter™?*

The mission of the Download Reporter™ is to provide industry summary statistics for computer software titles the same way the Neilson Ratings provide statistics for television shows and The New York Times provides information on best selling books.

These reports were originally developed for internal use at UDICo, because the current software market is much less forgiving than it was in the 80's and 90's.

- Depressed global economy means less spending
- More software titles in well defined categories means more competition
- Better hardware, operating systems and software means less reason to upgrade

There are more titles at a higher level of quality than ever before. What this means is that the marketplace has reached a point of diminishing returns, relative to its previous exponential growth.

In order to succeed in the current software publishing industry, quantitative and empirical methods must be applied to product definition and marketing.

The software industry is growing up – and we are growing up with it. While these reports were initially developed for in-house consumption, we hope that this information will be valuable to many software authors. As far as business intelligence goes, these reports are priced economically so that anyone who is interested in this information can reasonably afford it – if not on a subscription basis, then at least as a one time purchase.

## 1.2. *How to Use this Data*

Section 2 of this white paper has suggestions and examples of how to apply the data contained in these reports. The raw data by itself is interesting, but primarily it is a tool that can be used to answer questions about why strong titles have higher download counts than weaker titles.

In order to learn from the best and the worst companies, those companies first need to be identified. The Download Reporter™ supplies a statistical basis for categorizing software performance.

By understanding the behavior of the dominant companies within a category a startup can emulate their success factors. Likewise it is important to study the less successful companies and avoid their mistakes.

## 1.3. *Data Gathering Methodology*

The data structures used in the Download Reporter™ are derived from the information that is observed on Download.com. The following documentation describes the reports that are included in the report files. This information is combined with material from other Internet web sites.

## Limitations of Available Data

The information on CNet has certain inherent limitations.

- The Add Date identifies the CNet posting date for that record. It does not describe the publishing date for the title or the date the first version of that title appeared on Download.com.
- The Buy Price is an optional requirement and many publishers omit this information.
- The Total Count master pages contain titles that are no longer downloaded on a weekly basis.
- Some titles are over-represented by the inclusion of more than one release version.

The raw data is scrubbed before value based reports are generated.

## Two Primary Report Classes

The `Stat` table contains raw data pulled from CNet. This is the information exactly as it appears. The `StatValue` table is derived from that table and knocks out titles with no weekly activity for the previous six weeks.

Use the <code>Stat</code> table when you want the raw CNet data. Use the <code>StatValue</code> table when you want the scrubbed data.
--

Both data sets are included in the master report. The value of the raw data is that it contains unfiltered download counts and is the most primary source. The value of the `StatValue` table is that it is cleansed in order to correctly report estimates of economic value.

### 1.4. *Calculating Economic Value*

Potential economic value is at least as interesting as raw download count. In order to compare apples to apples the buy price and sale conversion rate must be considered with the raw download count – otherwise a shareware screensaver could appear to be more significant than an industry priced app server or software development tool.

The following is a high level description of the methodology used to construct the value based reports.

#### Determination of Buy Price

The buy price is the single most important field after weekly download count. All potential economic value is imputed from the buy price and download count.

- Software titles with no observed weekly download count and an `addDate` more than six weeks old are eliminated. This reduces the master data set from approximately 52,000 records to 48,000.
- Software titles with no `buyPrice` are researched on the web. In the case where a buy price is identified it is marked with a W in the `buyPriceInd` field.
- In the case where our staff has not yet found the price or no price is published, the industry average buy price for that title subcategory is used and the `buyPriceInd` field contains the letter A.

## Estimation of True Download Count

Each of the +30 aggregated markets are curved to create an expectation of download rates outside of CNet. For example, CNet is more important for some markets, such as Internet, than it is for other categories, such as IS/IT.

1. The weekly download count is multiplied by a market specific factor to determine the industry-wide download count for that title. For CNet-centric categories, such as Internet, a low multiplier of 5 is used to compute the total weekly download count. For other categories, such as Business & Finance, a higher multiple of 10 is used.
2. The Market Specific Factor is adjusted for Product Rank. Popular products are expected to have greater download counts than less popular products. Each Market is grouped by Download Count Quartile to select the final multiplier. For example, a first quartile Business & Finance product gets a download factor of 10, while a second quartile performer gets a download factor of 8.
3. The CNet weekly download count is multiplied by the Quartile Adjusted Market Specific factor to compute the cross-industry estimated weekly download count.

## Estimate of Try-to-Buy Conversion Rate

The try-to-buy conversion ratio determines the sales gross estimate that will be calculated for a title. Each of the +30 aggregated markets are curved to create an expectation of the buy-to-try conversion ratio. For example, some categories require more effort once the products are downloaded. In that case the assumption is that the download is more *well intentioned*, and that once the download decision is made the probability of a conversion is higher.

1. The estimated industry wide download count is multiplied by a market specific factor to estimate the unit sales for that title. For low-commitment markets such as Games, an average factor of 1.3% is used as the conversion ratio. For high-commitment markets, such as Business & Finance, an average value of 2.8% is used.
2. The Market Specific Factor is adjusted for Product Rank. Popular products are expected to have higher conversion ratios than less popular products. Each Market is grouped by Download Count Quartile to select the final multiplier. For example, a first quartile Business & Finance product gets a conversion ratio of 5%, while a second quartile performer gets a conversion ratio of 3.5%.
3. The product `buyPrice` is multiplied by the Quartile Adjusted Market Specific factor to compute the estimated gross sale per download.

## Estimated Weekly Sales Gross

The Estimated Weekly Sales Gross is calculated as the estimated gross revenue per download multiplied by the estimated cross-industry download count for that title.

## 2. Quantitative Marketing

### 2.1. *How to Use the Download Reporter™*

The statistics in the Download Reporter™ can be used to determine the factors that contribute to the success of some software titles and the failure of others. While the features that can be measured on Download.com represent a small subset of the universe of information regarding a product, they supply a basis for controlled and objective comparison and analysis.

### 2.2. *Have Realistic Expectations*

Don't expect to sell 20,000 copies of a product if the average weekly download count in that category is 100 units. Don't expect to price your product at \$1,000 if the CNet average is \$79 and the average in your category is less than \$200.

Use the statistics to form a baseline of anticipated product performance and economic value.

If nothing else, the Download Reporter™ reveals average performance within a category. While it is essential as an entrepreneur be optimistic, it is also important to plan for scenarios that encompass high, medium and low performance within an industry category.

### 2.3. *Listen to the Numbers*

1. Use Average and Hi/Lo performance to estimate scenarios for your products download count.
2. Knock out examples that skew the model. For example, the inclusion of mega-hits such as WinZip, ICQ or Morpheus will distort the numbers within a market.
3. Test your market assumptions by looking at additional performance data, such as 10K and Annual Report filings for publicly traded companies within your market sector.
4. Go to the websites of the best and worst titles within your group. Think about what the winners do that the losers do not.

### 2.4. *Try the Products and Form Your Own Opinion*

Download the best and worst of your competitors. Never forget the software publishing is about the entire customer experience.

### 2.5. *Do Your Own Content Analysis*

Carefully read the titles, short and long form descriptions of the best and worst products in your category. Think about the features that are stressed by the strong products, versus those that are mentioned by the weak ones. Use statistical tools like histograms of word usage to discover key words, phrases and product features. Verify that your products looks and sounds like a winner.



## 2.6. ***Look for Correlated Success Factors***

What do you think about when you look for a product on Download.com?

1. Is the software popular? What is the weekly download count?
2. Is the software fresh? When was the title last updated?
3. How many reviews does the title have and what do they say?
4. Is there a fully functional free evaluation version?
5. What is the buy price, especially versus other titles in the same group?
6. What is the file size relative to the average in that group? Less is more, and what may be OK for an HTML Editor might be unacceptable for a utility.

These factors represent a sample of the product features that may be important in your software category. Look for correlations that tend to separate the high performing titles from the low performing ones.

Our title level detail report includes the product description, so a fair amount of content analysis is possible using that report as a starting point. Again, once you identify the winners and losers, it is possible to do a marketing analysis to find the correlated factors that push them apart. The product page URL is also included in the Category Detail reports, so its possible to drill down from the Download Reporter™ directly to the CNet product page.

## 2.7. ***Putting it All Together***

### **Choose Your Market**

Look at Your Product and decide what categories it could possibly fit in - then develop an opinion of expected success rate, i.e., download count, based upon averages and hi/lo's in those categories.

### **Find Your Competitors**

Next search for the titles in those categories that are closest in product type to Your Product.

### **Sound Like a Winner**

Now do a content analysis of those products to deduce the important keywords or phrases.

### **Identify the Success Factors**

Do not fail to include a correlation to the other factors that make the winners successful, i.e., price, file size, time at market, quality of product, etc.

## 3. Overview of Data Structures

### 3.1. Reporting Tables Summary

#### The Stat Table

The `Stat` or *Statistics* table holds basic indicative information for the software titles. This information is compiled from observations of category master pages and product title pages.

#### Category Master Sample

```
http://download.com.com/3150-2150-0-1-5.html
```

The Category Master pages contain statistics about collections of titles.

#### Product Page Sample

```
http://download.com.com/3000-2150-7209042.html
```

Product pages contain detail on the individual titles.

### 3.2. Notes on Selected Fields

The following notes clarify the content of key fields used in multiple reports.

#### **url**

The `url` field does not contain `http://www.download.com.com`. Note that the `.com.com` is not a typo, but is the correct URL prefix. For example:

```
/3000-2150-7209042.html
```

This URL becomes:

```
http://download.com.com/3000-2150-7209042.html
```

as the valid CNet URL.

#### **category**

The CNet category number. Note that in aggregated reports the number 1000 is used. For example, category #2029 is Linux Development Tools, while 1000 is All CNet as well as other aggregated categories such as Windows Games.

#### **pubName, pubUrl**

This information is retrieved by clicking through the category listing. In cases where the page is down, the publisher information may not be available for the report.

**fsize**

The file size is given in KB. In the case of aggregated reports, this is the average value for that group.

**tryPrice**

There is only one company with a try price listed on CNet, and this is a typo – they meant to list the `buyPrice`. This field may be dropped from subsequent report versions.

**buyPrice**

Not every company lists a `buyPrice`. Our research staff is currently working its way through all of the titles to discover this information empirically. However, there are some cases where the software does not have a `buyPrice` and the software is actually free or the revenue model is driven by advertising or other means. In this case the CNet average buy price for that sub category will be used to construct value based reports. In some cases, if software is discovered to have no commercial value, it may be excluded from value based reports.

**buyPriceInd**

If the `buyPriceInd` is not set, then the value was observed on CNet. If the value is set to W it has been researched on the web. If the value is A it is taken from the industry average for that software category.

**addDate**

The CNet add date for that record. This is not the same as the age of the software title, and may not even correspond the publishing date for that product version. The `addDate` is, however, the age of the CNet listing and can be used to derive some information about the *staying power* of a title.

**voteInd**

This field contains `Thumbs Up` or presumably `Thumbs Down`, although the only observed values are `Thumbs Up` and blank.

**votePnt**

Percent of votes corresponding to the `voteInd` field.

**voteCnt**

The number of votes recorded for that title. The average number for aggregated reports.

**weeklyCnt**

The CNet weekly download count.

**totalCnt**

The CNet total download count.

### 3.3. *Title Rank, Quartile and Percentiles*

All values in the **stat** report (reports by count) are based on the **Weekly Download Count**. All values in the **statValue** table are based on the **Estimated Weekly Sales Gross** for that title.

#### **opsQuartile**

The cross-operating system rank by quartile where 1 is the best.

#### **opsPercentile**

The cross-operating system rank by percentile where 99 is the best. Note that in cases where there are a limited number of titles in a group, the best score will be less than 99.

#### **opsRank**

Absolute rank relative to all titles in that operating system.

#### **mktQuartile**

The cross-market system rank by quartile where 1 is the best.

#### **mktPercentile**

The cross-market system rank by percentile where 99 is the best. Note that in cases where there are a limited number of titles in a group, the best score will be less than 99.

#### **mktRank**

Absolute rank relative to all titles in that product market.

### 3.4. *Ranks by Category and Subcategory*

These measurements follow the model used for operating system and market. Note that not every CNet market is subdivided into categories and subcategories, so the values listed in those columns may be zero.

## 4. The Stat Table Data Dictionary

Stat Table		
Name	Data Type	Description
url	VarChar(48)	CNet URL with download.com omitted.
inpDate	Datetime	Capture date of record.
chgDate	Datetime	Change date of record if modified after initial capture.
category	Char(4)	CNet 4 digit software category.
pubName	VarChar(64)	Software publisher name.
pubUrl	VarChar(128)	Software publisher URL.
title	VarChar(128)	Software product title.
descr	VarChar(255)	Short form product description.
opsys	VarChar(32)	Operating system.
fsize	Number	File size in KB.
sponsFlg	Bool	Sponsored Link 1 if TRUE.
tryPrice	Number	Try price if published.
buyPrice	Number	Buy price if available. If the buyPrice is not listed on CNet, then the publisher web site is consulted.
buyPriceInd	Char(1)	Blank if from CNet, W=From Web, A=Category Average.
addDate	Datetime	CNet title addDate - Note that this does not always correspond to the software introduction data or even the software version date.
voteInd	VarChar(12)	Thumbs Up or Thumbs Down. Note that only Thumbs Up has been observed.
votePnt	Number	Vote Percent - Percentage positive or negative depending upon voteInd.
voteCnt	Integer	Vote Count -The number of votes recorded.
weeklyCnt	Integer	Weekly Count - The weekly number of downloads reported on CNet.
totalCnt	Integer	Total Count - The total number of downloads recorded on CNet.
opsQuartile	Number	Quartile score by OS where 1 is the best.
opsPercentile	Number	Percentile score by OS where 99 is the best. Note that in categories with a low number of examples the highest score will be less than 99.
opsRank	Number	Absolute Rank by OS where 1 is the best.
mktQuartile	Number	Quartile score by Market where 1 is the best.
mktPercentile	Number	Percentile score by Market where 99 is the best. Note that in categories with a low number of examples the highest score will be less than 99.
mktRank	Number	Absolute Rank by Market where 1 is the best.
catQuartile	Number	Quartile score by Category where 1 is the best. Note that not all Markets are subdivided into separate categories.
catPercentile	Number	Percentile score by Category where 99 is the best. Note that in categories with a low number of examples the highest score will be less than 99. Note that in cases where there is no category breakdown this number will be zero.
catRank	Number	Absolute Rank by Category where 1 is the best.
subQuartile	Number	Quartile score by SubCategory where 1 is the best. Note that not all Categories are subdivided into separate SubCategories.
subPercentile	Number	Percentile score by SubCategory where 99 is the best. Note that in categories with a low number of examples the highest score will be less than 99. Note that in cases where there is no category breakdown this number will be zero.
subRank	Number	Absolute Rank by SubCategory where 1 is the best.

## 5. The StatValue Table Data Dictionary

Stat Table		
Name	Data Type	Description
url	VarChar(48)	CNet URL with download.com omitted.
inpDate	Datetime	Capture date of record.
category	Char(4)	CNet 4 digit software category.
pubName	VarChar(64)	Software publisher name.
pubUrl	VarChar(128)	Software publisher URL.
title	VarChar(128)	Software product title.
descr	VarChar(255)	Short form product description.
opsys	VarChar(32)	Operating system.
sponsFlg	Bool	Sponsored Link 1 if TRUE.
buyPrice	Number	Buy price if available. If the buyPrice is not listed on CNet, then the publisher web site is consulted.
buyPriceInd	Char(1)	Blank if from CNet, W=From Web, A=Category Average.
weeklyValue	Number	A formula is used to calculate the estimated weekly value. See additional documentation.
totalValue	Number	A formula is used to calculate the estimated total value. See additional documentation.
addDate	Datetime	CNet title addDate - Note that this does not always correspond to the software introduction data or even the software version date.
dayCnt	Integer	Days between addDate and inpDate.
weekCnt	Integer	Weeks between addDate and inpDate.
weeklyCnt	Integer	Weekly Count - The weekly number of downloads reported on CNet.
avgWeeklyCnt	Integer	The totalCnt divided by the weekCnt. Note this is not always a valid indicator of change in download rate.
totalCnt	Integer	Total Count - The total number of downloads recorded on CNet.
opsQuartile	Number	Quartile score by OS where 1 is the best.
opsPercentile	Number	Percentile score by OS where 99 is the best. Note that in categories with a low number of examples the highest score will be less than 99.
opsRank	Number	Absolute Rank by OS where 1 is the best.
mktQuartile	Number	Quartile score by Market where 1 is the best.
mktPercentile	Number	Percentile score by Market where 99 is the best. Note that in categories with a low number of examples the highest score will be less than 99.
mktRank	Number	Absolute Rank by Market where 1 is the best.
catQuartile	Number	Quartile score by Category where 1 is the best. Note that not all Markets are subdivided into separate categories.
catPercentile	Number	Percentile score by Category where 99 is the best. Note that in categories with a low number of examples the highest score will be less than 99. Note that in cases where there is no category breakdown this number will be zero.
catRank	Number	Absolute Rank by Category where 1 is the best.
subQuartile	Number	Quartile score by SubCategory where 1 is the best. Note that not all Categories are subdivided into separate SubCategories.
subPercentile	Number	Percentile score by SubCategory where 99 is the best. Note that in categories with a low number of examples the highest score will be less than 99. Note that in cases where there is no category breakdown this number will be zero.
subRank	Number	Absolute Rank by SubCategory where 1 is the best.

## 6. The CategoryReport Table

Category Report Table		
Name	DataType	Description
id	Char(4)	CNet 4 digit software category.
os	VarChar(32)	Operating system, e.g., Windows
market	VarChar(48)	Software Market, e.g., Software Development
category	VarChar(48)	Software Category, e.g., Java
subCategory	VarChar(48)	Software Subcategory, e.g., Source Code
fsize	Number	Average file size in KB.
sponsFlg	Number	Total number of sponsored pages in group.
tryPrice	Number	Average try price in group.
tryPriceCnt	Number	Number of titles with try price listed.
buyPrice	Number	Average buy price of those listed.
buyPriceCnt	Number	Number of titles with Buy Price Listed in group.
voteInd	Number	Constant value 'Thumbs Up'.
votePnt	Number	Average vote percent in group.
voteCnt	Number	Total number of votes in group.
voteCntAvg	Number	Average number of votes per title.
weeklyCnt	Number	Weekly downloads in group.
weeklyCntAvg	Number	Average weekly download count for group.
totalCnt	Number	Total number of downloads in group.
totalCntAvg	Number	Average number of downloads per title.
groupCnt	Number	Number of titles in group.

## 7. The CategoryValue Table

Category Value Table		
Name	Data Type	Description
id	Char(4)	CNet 4 digit software category.
os	VarChar(32)	Operating system, e.g., Windows
market	VarChar(48)	Software Market, e.g., Software Development
category	VarChar(48)	Software Category, e.g., Java
subCategory	VarChar(48)	Software Subcategory, e.g., Source Code
sponsFlg	Number	Total number of sponsored pages in group.
buyPrice	Number	Average buy price of those listed.
dayCnt	Number	Number of days since addDate.
weeklyValue	Number	Estimated weekly sales gross.
weeklyValueAvg	Number	Estimated average gross per title.
totalValue	Number	Total estimated sales gross since addDate.
totalValueAvg	Number	Total average sales gross since addDate.
weeklyCnt	Number	Weekly downloads in group.
weeklyCntAvg	Number	Average weekly download count for group.
totalCnt	Number	Total number of downloads in group.
totalCntAvg	Number	Average number of downloads per title.
groupCnt	Number	Number of titles in group.



## 8. UDICo Contact Information

### 8.1. *Company Overview*

The Universal Data Interface Corporation is a technology spin-off from PricewaterhouseCoopers LLP. The founder of the company, Adam Greissman, is specialist in designing and implementing globally distributed transaction processing systems and a pioneer in the development of XML specifications for financial transactions. The Universal Data Interface is a middleware product targeted at EAI and B2B applications, and has been used to implement systems for financial institutions, beginning with the Chase Manhattan Bank in 1993. The UDI has been deployed for applications such as front-to-back office reconciliation, credit and market risk management (Sungard and Reuters), ERP integration for SAP and PeopleSoft, and CRM applications such as BroadVision and Siebel. The UDI middleware engine is the foundation of TierBroker.

### 8.2. *Contact Information*

Questions about this document may be addressed to the following UDICo officers:

**Adam Greissman**  
**Chief Executive Officer**  
917.497.4342 – alg@UDICo.com

**Universal Data Interface Corporation**  
80 Wall Street, Suite 215  
New York, NY 10005