Web Aggregators and Aggregation Research:
Prof. Stuart Madnick
(smadnick@mit.edu)

© S. Madnick & M. Siegel, 2005

Multiple Perspectives

Strategy and Policy

Application

Aggregators

Global Financial Account Aggregation

Web wrapper & Context mediator

Technology

eCommerce & Financial Services
Overview: Aggregation – Focus on Information Re-use and Re-purposing

- Aggregator definition, types & examples
- Emerging Global Aggregation
- Enabling Technologies
- New Emerging Aggregators
- Strategy, Legal, Cultural, and Policy Issues
- Summary

Aggregator Definition

- Transparently collects information
  - with or without prior source arrangement
  - resolves contextual differences
- Performs analysis
  - not merely gather
- Represents user
  - acts as agent for user
## Aggregator Types & Examples

<table>
<thead>
<tr>
<th>Types</th>
<th>Internal</th>
<th>External</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparison</strong></td>
<td><em>Benchmarking divisions</em></td>
<td><em>Shop for best prices (e.g., books)</em></td>
</tr>
<tr>
<td><strong>Relationship</strong></td>
<td><em>Customer Relationship Management (CRM)</em></td>
<td><em>Financial account aggregation</em></td>
</tr>
<tr>
<td><strong>Complementary</strong></td>
<td><em>Department calendars</em></td>
<td><em>Investment analysis, policy analysis</em></td>
</tr>
</tbody>
</table>

---

### Sources

**Internal**
- Benchmarking divisions
- Customer Relationship Management (CRM)
- Department calendars

**External**
- Shop for best prices (e.g., books)
- Financial account aggregation
- Investment analysis, policy analysis

---

**www.amazon.com**

*Microsoft Secrets: How the World’s Most Powerful Software Company Creates Technology, Shapes Markets, and Manages People*

- **by Michael A. Cusumano, Richard W. Seltzer (Contributor)**
- **List Price**: $21.00
- **Our Price**: $15.75 (20%)
- **You Save**: $5.25 (30%)

**Availability**: Usually ships within 24 hours.

**Handcover** (October 1995)
- **Format**: Hardcover
- **Dimensions**: 8.5 x 5.5 x 1.17 in.
- **Weight**: 1.57 lb
- **Other Editions**: Paperback

**Amazon.com Sales Rank**: 32,674

**Avg. Customer Review**: 80% of 939
- **Number of Reviews**: 939

Write an online review and share your thoughts with other readers!

Customers who bought this book also bought:
Microsoft Secrets: How The World's Most Powerful Software Company Creates Technology, Shapes Markets, And Manages People

Today, Microsoft commands the high ground of the information superhighway by owning the operating systems and basic applications programs that run on the world's 370 million computers. Beyond the unquestioned genius and vision of Bill Gates, what accounts for Microsoft's astounding success?

For the first time, drawing on almost two years of on-site observation at Microsoft

* Acquired by Barnes & Noble
Mega-Aggregators

* Note: multiple book price aggregators
  - using differing sources
  - giving differing results!

* What about an aggregator of aggregators
  - A Mega-Aggregator ... more later ...

* Aggregators not just about prices ...

---

www.evenbetter.com *  
Now called DealTime.com  
Comment: Need for Mega-Aggregators
### www.usair.com

**US AIRWAYS DIVIDEND MILES**

Current Dividend Miles Account Statement

- **Member Name:** DR. STUART E MADNICK
- **Dividend Miles Account Number:** 561316
- **Membership Level:** Dividend Miles Member
- **Miles Posted as of:** July 22, 1999

**Account Statement**

- **Current Balance:** 36,216
- **Year-To-Date Preferred Miles:** 2,000
- **Year-To-Date Preferred Segments:** 4

**Account Activity Since Your Last Statement**

<table>
<thead>
<tr>
<th>Date</th>
<th>Transaction Description</th>
<th>Actual Miles</th>
<th>Bonus Miles</th>
<th>Total Miles</th>
</tr>
</thead>
</table>

### www.maxmiles.com

**3 Miles/MDM Sample Report - Microsoft Internet Explorer provided by America Online**

**Account Summary**

<table>
<thead>
<tr>
<th>Program</th>
<th>Account Balance</th>
<th>Program Status</th>
<th>Account Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avianca</td>
<td>70,000</td>
<td>Member</td>
<td>200,000</td>
</tr>
<tr>
<td>Continental</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>Delta</td>
<td>7,754</td>
<td>Silver Member</td>
<td>0</td>
</tr>
<tr>
<td>Hilton</td>
<td>3,070</td>
<td>Member</td>
<td>600</td>
</tr>
<tr>
<td>Hyatt</td>
<td>34,313</td>
<td>Platinum</td>
<td>714</td>
</tr>
<tr>
<td>Mayo</td>
<td>6,590</td>
<td>Red</td>
<td>807</td>
</tr>
<tr>
<td>No Almost</td>
<td>6,470</td>
<td>Gold</td>
<td>300</td>
</tr>
<tr>
<td>United</td>
<td>7,611</td>
<td>Premier Executive</td>
<td>807</td>
</tr>
<tr>
<td>M德尔 Expiry</td>
<td>500</td>
<td>Member</td>
<td>100</td>
</tr>
</tbody>
</table>

**Expiration Summary**

<table>
<thead>
<tr>
<th>Program</th>
<th>No Expiration</th>
<th>Begin Date</th>
<th>End Date</th>
<th>Begin Miles</th>
<th>End Miles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avianca</td>
<td>0</td>
<td>7,400</td>
<td>3,000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Continental</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Delta</td>
<td>7,754</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hilton</td>
<td>3,070</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hyatt</td>
<td>34,313</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mayo</td>
<td>6,590</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No Almost</td>
<td>6,470</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>United</td>
<td>7,611</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
What is MaxMiles

You give it all your frequent flyer info, it . . .

- Does mileage management for you . . .
- Gathers all your frequent flyer info
- Sends a personalized report via email
- Updates reports on web site daily
- Searches for missing mileage credits
- Analyzes most current milage offers

From US Airways site

"Access to this information is subject to the rules in the Dividend Miles Membership Guide. In addition, by using this site to access your Dividend Miles account, you agree that you will use this site in a manner consistent with the Dividend Miles Membership Guide and you further agree not to allow access to this site to any third party by revealing your access code to any third party for any reason. Failure to comply with the foregoing restrictions on the use of this site shall be grounds for the termination of your access to this site and/or your membership in the Dividend Miles program."

“Power of Attorney” on MaxMiles Site

**Limited Power of Attorney**

By clicking in this checkbox, I am indicating that I have read and that I accept the following terms and grant MaxMiles, Inc. (MaxMiles) the rights described below:

I authorize MaxMiles to access and obtain my frequent flyer account information from the frequency programs described on this registration form, as the same may be amended from time to time. For all purposes hereof, MaxMiles is granted a limited power of attorney, and I hereby appoint MaxMiles as my true and lawful attorney-in-fact and agent, with full power of substitution and resubstitution, for me and in my name, place and stead, in any and all capacities, to access such data from the frequency programs at all times so MaxMiles may determine in its sole discretion, with the full power and authority to do and perform each and every act and thing requisite and necessary to be done in connection with such access, as fully to all intents and purposes as I could do in person. MaxMiles is further authorized to utilize such information in preparing reports regarding my accounts and can consolidate this information for MaxMiles use as it determines in its sole discretion. I understand that MaxMiles acts solely as an intermediary, that my benefits are subject to the program sponsor’s individual terms and restrictions, and that MaxMiles has no authority to access, direct or divert any program award or other benefits. I further understand that MaxMiles has no relationship to or connection with any airline, hotel, or other frequent traveler program sponsor, and agree that MaxMiles assumes no responsibility and shall incur no liability with respect to the acts, omissions or determinations of any program sponsor.

**Complementary Aggregator**

**Integrated Spreadsheet**

**Sources**

- Real time
- Tibco

**External**

**Internal**
Everyone is impacted:
If you are not an Aggregator,
you are an Aggregatee

- Books.com, MySimon.com, EvenBetter.com
  - Barnes and Noble, Borders, Amazon, ...
- MaxMiles
  - United, American, TWA, US Air, ...
- Universal Financial Aggregator (UFA)
  - your banks, your brokers, your billers, ...

• NOTE: Aggregatees -> Aggregators

Overview

- Aggregator definition, types & examples
- Emerging Global Aggregation
- Enabling Technologies
- New Emerging Aggregators
- Strategy, Legal, Cultural, and Policy Issues
- Summary
Global Comparison Aggregator

Want to find the best online price?

<table>
<thead>
<tr>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regional Aggregators</td>
</tr>
<tr>
<td>Price Dispersion</td>
</tr>
<tr>
<td>Price? What is it really?</td>
</tr>
</tbody>
</table>

Are you interested in shopping for a SONY DCR-UPS, a pocket-sized digital camcorder that weighs only 12 oz (367g), records in MPEG format, and looks like the one shown here?

Yes!

OK, we will help you find the best prices online! Which country are you from or would you like to do your shopping?

<table>
<thead>
<tr>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
</tr>
<tr>
<td>France</td>
</tr>
<tr>
<td>Italy</td>
</tr>
<tr>
<td>Japan</td>
</tr>
<tr>
<td>Mexico</td>
</tr>
<tr>
<td>The Netherlands</td>
</tr>
<tr>
<td>Norway</td>
</tr>
<tr>
<td>Spain</td>
</tr>
<tr>
<td>Sweden</td>
</tr>
<tr>
<td>UK</td>
</tr>
<tr>
<td>USA</td>
</tr>
</tbody>
</table>

http://soursop.mit.edu/global

Best global online price

Best Prices Found …

<table>
<thead>
<tr>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regional Aggregators</td>
</tr>
<tr>
<td>Price Dispersion</td>
</tr>
<tr>
<td>Price? What is it really?</td>
</tr>
</tbody>
</table>

We searched 16 online stores in Sweden and the lowest price was $1760.

Even better.

We also searched stores outside Sweden. We searched a total of 172 stores worldwide and the lowest price was $783 from a store in The Netherlands.

In case you are interested, the highest price was $2237 from a store in Sweden.

Is that cool to find out the world best prices when you shop online? Out of curiosity, you may have a few questions, such as:

- How prices of the camcorder differ in various parts of the world?
- Take a guess. Then go to Price Dispersion for some hint.

- Is there a global bargain finder like this for me to use?
- Not really. Although there are tools that search prices regionally.

- Does the lowest price from the Netherlands include shipping?
- Great question! See “Price? What is it really?” for explanations.

- What are the challenges for building a global bargain finder?
- See the Design section for our approach and a live demonstration.
Global Complementary Aggregation

Q: How did CO2 emissions (total, per GDP, per capita) change over time (between 1990 and 2000) in Yugoslavia?

- User 1: YUG as a geographic region bounded before the breakup
- User 2: YUG as a legal autonomous state

Related effort:
- Laboratory for Information Globalization and Harmonization Technologies (LIGHT)

<table>
<thead>
<tr>
<th>Country</th>
<th>Code</th>
<th>Currency</th>
<th>Code</th>
<th>Currency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yugoslavia</td>
<td>YUG</td>
<td>New Yugosl. Dinar</td>
<td>YUG</td>
<td>YUN</td>
</tr>
<tr>
<td>Bosnia and Herzegovia</td>
<td>BIH</td>
<td>Marka</td>
<td>BAM</td>
<td></td>
</tr>
<tr>
<td>Croatia</td>
<td>HRV</td>
<td>Kuna</td>
<td>HRK</td>
<td></td>
</tr>
<tr>
<td>Macedonia</td>
<td>MKD</td>
<td>Denar</td>
<td>MKD</td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td>SVN</td>
<td>Tolar</td>
<td>SIT</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>YUG</td>
<td>35604</td>
<td>15480</td>
<td>1279</td>
<td>5405</td>
<td>3378</td>
<td>3981</td>
</tr>
<tr>
<td>BIH</td>
<td>1279</td>
<td>5405</td>
<td>3378</td>
<td>3981</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRV</td>
<td>5405</td>
<td>3378</td>
<td>3981</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MKD</td>
<td>3378</td>
<td>3981</td>
<td>3981</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVN</td>
<td>3981</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Many sources needed:
Meanings in sources & users might differ
Overview

- Aggregator definition, types & examples
- Emerging Global Aggregation
- Enabling Technologies
- New Emerging Aggregators
- Strategy, Legal, Cultural, and Policy Issues
- Summary

Emerging Technologies

- Web wrappers
  - Extract selected information from web
  - Allows web to be treated as large database

- Context mediators
  - Resolve semantic (meaning) differences
  - Enable meaningful aggregation
    & comparison
Overview

- Aggregator definition, types & examples
- Emerging Global Aggregation
- Enabling Technologies
- New Emerging Aggregators
- Strategy, Legal, Cultural, and Policy Issues
- Summary

Concept of a Universal Financial Aggregator (UFA)

User

UFA

Citigroup
Credit Cards

Fidelity
Brokerage

Fleet
Banking

BancOne
Credit Cards
COntext INterchange Project (context2.mit.edu/coin)

Universal Financial Aggregator (UFA) Demonstration
**Aggregated Report produced (top part)**

**Deposit Accounts**

<table>
<thead>
<tr>
<th>Bank Name</th>
<th>Account Number</th>
<th>Nickname</th>
<th>Account Type</th>
<th>Available Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>BankOne</td>
<td>408905465</td>
<td>234234</td>
<td>Checking</td>
<td>220.00</td>
</tr>
<tr>
<td>Security First</td>
<td>12399662</td>
<td>My general</td>
<td>MMA</td>
<td>200.00</td>
</tr>
<tr>
<td>Network Bank</td>
<td></td>
<td>account</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citibank</td>
<td>78019470</td>
<td>School 1</td>
<td>Checking</td>
<td>987.00</td>
</tr>
</tbody>
</table>

Total: **5157.00**

**Credit Card Accounts**

<table>
<thead>
<tr>
<th>Issuer</th>
<th>Account Number</th>
<th>Nickname</th>
<th>Available Credit</th>
<th>Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>FirstUSA</td>
<td>4137416890</td>
<td>Misc. Card</td>
<td>788.00</td>
<td>1212.00</td>
</tr>
<tr>
<td>FirstUSA</td>
<td>31245690</td>
<td>Emergency Card</td>
<td>900.00</td>
<td>1320.00</td>
</tr>
</tbody>
</table>

Total: **1610.00**

**Aggregated Report produced (bottom part)**

**Brokerage Accounts**

<table>
<thead>
<tr>
<th>Broker</th>
<th>Account Number</th>
<th>Account Name</th>
<th>Nick Name</th>
<th>Total Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>E*TRADE</td>
<td>90356</td>
<td>SASHAM</td>
<td>Daybeam Account</td>
<td>944.08.00</td>
</tr>
<tr>
<td>E*TRADE</td>
<td>743241</td>
<td>Steven</td>
<td>Trading Account</td>
<td>3812.75</td>
</tr>
<tr>
<td>E*TRADE</td>
<td>545631</td>
<td>test1</td>
<td>test1</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Total: **195781.50**

**Net Worth**

<table>
<thead>
<tr>
<th>Assets</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Shares</td>
<td></td>
<td></td>
<td></td>
<td>9950.00</td>
</tr>
<tr>
<td>Brokerage</td>
<td></td>
<td></td>
<td></td>
<td>10010.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Liabilities</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit Card Accounts</td>
<td></td>
<td></td>
<td></td>
<td>1312.00</td>
</tr>
</tbody>
</table>

Total: **156026.50**
“Aggregation . . . The ‘killer app’ in the e-finance revolution”

Current UFA Offerings (examples)

Account Aggregation Technology Suppliers
• Yodlee
• VerticalOne
• eBalance

Account Aggregation Offerings
• Chase, Fleet, Citibank (myCiti.com), . . .
• Merrill-Lynch, Fidelity Investments, . . .
• Yahoo, AOL, . . .
Commercial Financial Account Aggregators

Real Value: “After Aggregation” Services

- First Call
- User Information
- Account Analyst
  - Advisors
  - Wizards
  - Agents

Other Internet Sources
- Financial Sources
- Intranet Sources
  - Other Internet Sources
  - Intranet Sources and applications

Excel
Advisors: Stock holdings re-aggregate

Wizards: Money Market (MMA) example
MMA is also an aggregator, of MM rate sources

The Universal Financial Aggregator (UFA):
- Aggregates your financial information
  - several ways: totals and stock holdings
- Also uses the Money Market Aggregator
  - which draws on multiple MM sources
  - including sources that are, in fact, aggregators of some MM sources
UFA Goals

- Consumers and Financial Institutions:
  - Can do what they do better
  - Can do things they could not do before
  - Redefine the use, role, and relationship of financial service
- Business-to-Business examples also

Overview

- Aggregator definition, types & examples
- Emerging Global Aggregation
- Enabling Technologies
- New Emerging Aggregators
- Strategy, Legal, Cultural, and Policy Issues
- Summary
Strategy, Legal, Cultural, and Policy Issues Regarding Aggregation

- Relationships between Aggregators and Aggregatees
- Protection of investment in data vs re-use
  - EU Database Directive
  - 4 proposed bills in US
- Protection of consumer privacy
- Liability, consumer protection, jurisdiction
Aggregator “states” example

Intershipper

Independent Agent (BITS, Inc.)

Independent Aggregator

Partial Collaborative Aggregator

iShip

Dependent Aggregator / Dominant (UPS)

Dependent Aggregator / Consortium
Database Re-Use Lawsuits and Claims

**Example Cases**
- eBay vs. Bidder’s Edge
- HomeStore.com vs. Bargain Network
- First Union vs. PayTrust
- mySimon vs. Priceman
- Register.com vs. Verio
- Ticketmaster vs. Tickets.com

**Principles Used**
- Copyright infringement
- Trespass to chattels
- Breach of contract
- Misappropriation
- False advertising
- Violation of the Computer Fraud and Abuse Act

“Copyright infringement”
- Protects original works of authorship
- Usually not applicable to data re-use

“Trespass to chattels”
- eBay vs. Bidder’s Edge
  - Injunction on May 24, 2000
- Ticketmaster vs. Tickets.com
  - “It is hard to see how entering a publicly available web site could be called a trespass”
Battle over Database Re-Use - continued

- **New Paradigm – “sweat of the brow”**
  - Collections of Information Antipiracy Act, H.R.354 (1999) – Coble bill
  - Two other failed bills in 1996, 1997
  - EU Database Directive

International Differences

**U.S.**
- Decentralized
- Self-regulation
- No DB protection law
- Opt-out for privacy

**E.U.**
- Centralized
- Specific, strict regulation
- DB Directive
- Opt-in for privacy

- **Reciprocity provision**
  - Not protected if home country does not have similar law for DB protection
  - No data flow to countries without similar privacy protection law

- **Harmonization is needed**
  - Safe Harbor Agreement between US and EU
Overview

- Aggregator definition, types & examples
- Emerging Global Aggregation
- Enabling Technologies
- New Emerging Aggregators
- Strategy, Legal, Cultural, and Policy Issues
- Summary

Key Global Aggregation Insights & Implications

<table>
<thead>
<tr>
<th>Insights</th>
<th>Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Everyone impacted</td>
<td>Be part of strategy planning</td>
</tr>
<tr>
<td>Aggs not limited to price</td>
<td>Identify new Agg applications and business opportunities</td>
</tr>
<tr>
<td>comparison</td>
<td></td>
</tr>
<tr>
<td>Aggs can be combined and</td>
<td>Entirely new types of applications can be defined</td>
</tr>
<tr>
<td>aggregated</td>
<td></td>
</tr>
<tr>
<td>Aggs create new info spaces</td>
<td>Who will own this space</td>
</tr>
<tr>
<td>After-agg even more important</td>
<td>This must be exploited</td>
</tr>
<tr>
<td>New technologies emerging</td>
<td>Exploit these technologies</td>
</tr>
<tr>
<td>Business models and market</td>
<td>Must be prepared for changes</td>
</tr>
<tr>
<td>conditions continue to change</td>
<td></td>
</tr>
</tbody>
</table>
Current Related Projects (partial list)

- Aggregation strategies (e.g., Open aggregation)
- Regional & Global Aggregation: Cultural & legal
- Multi-enterprise eBusiness: Trust & relationships
- Financial services/Internet support for SME’s
- Future Financial Services (e.g., Wireless aggregator)
- Context Mediation & Web data extraction *
- Data Quality & What XML will and will not do
- Internal & Industry data standards efforts
- Other new technologies (e.g., Mobile aggregation)

* Patents awarded in 1999

For Further Information and Contacts

- MIT Research:
  - http://context2.mit.edu and http://context2.mit.edu/aggregation
  - http://web.mit.edu/tdqm

- MIT Technology Licensed Software:

- Email Contact:
  - Stuart Madnick at smadnick@mit.edu
  - Michael Siegel at msiegel@mit.edu

Research funded, in part, by Fleet Bank, Merrill Lynch, Suruga Bank, Banco Santander (BSCH), Citibank, DoD Advanced Projects Agency (DARPA), MITRE Corporation, PricewaterhouseCoopers, Singapore-MIT Alliance (SMA), MIT-Malaysia University of Science and Technology (MUST) Collaboration