



The Drive for Content Personalization

by Mark Daniel and Lee McCaleb

EMC Corporation

Contents

What is content personalization?	3
How to define a consumer's content preferences	4
Added benefits of a content repository managed by EMC Documentum	6
No database, no service	6
The need for content context is universal	9

The Drive for Content Personalization

Most of us have read about the future of media—how we’ll watch what we want to watch on any device we have at hand. We’ll see commercial messages that have been customized to our unique interests and perhaps even see programs personalized to those same preferences.

Considering what we have experienced over the decades from mainstream media, that kind of change is more than significant; it’s revolutionary. The classic media industry model is based on offering content that appeals to the majority of people, to broadly cast (and that’s no play on words) that content to the masses. Remember the term “mass media?” How does mass media become “personal media” without splitting itself apart?

This change in the way we consume content is fueled by the ubiquitous acceptance of the Internet and its ability to convey information and content that is infinitesimally granular. It’s great to be able to watch videos of kids dropping Mentos mints into Diet Coke bottles to create the now-famous “gusher” reaction, but it’s doubtful that anyone has made a living actually **creating** content of this sort. The 2006 *Time* magazine “Person of the Year” was declared to be all of us, citizens of the digital democracy changing the information and entertainment worlds for a lark, not for a buck.

Major media companies can learn from these Web content examples. But they also need to be cautious as they move away from the business models that have built their businesses for many decades into the new world of targeted content and one-on-one consumer relationships.

What is content personalization?

Assume that people will one day be able to watch what they want, on any device, at any time. That doesn’t mean they know what they want to watch at any given time. Infinite choice can lead to confusion if people are confronted with too much that is unfamiliar to them. How do you know you will like a program if you have never heard of it before, and it’s performed by artists with whom you are unfamiliar? You don’t know. And that’s why most people stick with what is familiar and has proven enjoyable in the past.

Content personalization is nothing more than focusing on what consumers prefer and making sure that type of content is brought to their attention. To do that, one must define a set of preferences for each consumer. But people are extremely complex. It’s virtually impossible to find a person who “only likes comedy” and will watch nothing else on their television. People’s tastes run the gamut, so how do you know what to offer?

You certainly cannot count on people knowing their own preferences. People claim to like certain types of programs, but never actually spend time watching them. Nielsen Media Research discovered this fact while creating their first rating system for audience size in broadcasting. Originally, Nielsen asked people to fill-out diaries stating what they watched on which channels at which times. Nielsen found that people reported what they wanted other people to **think** that they were watching, not what they actually watched. Grown men did not want to admit that they were watching cartoons and not the evening news at 5:30 p.m.

Real content-preference information comes from tracking what people actually watch. In a time of video-on-demand (VOD) services, this information becomes even more comprehensive, accurate, and useful because people are choosing from potentially thousands of programs, some of which the consumers must pay an additional charge beyond their normal monthly bill. From this insight, it may be possible to see viewing-habit characteristics and trends that ultimately allow the content provider to extrapolate that consumer’s preferences.

How to define a consumer's content preferences

Many content providers believe a process known as “activity tracking” is beginning to give them a way to track consumer preferences by tracing the consumer's choices as interactive transactions at all times.

Activity without context, however, is not real information. What does it mean when a person looks into a shop window when you don't know which merchandise caught their eye? Bertrand Russell, the British philosopher, said, “There are two motives for reading a book: One, that you enjoy it, and the other, that you can boast about it.”

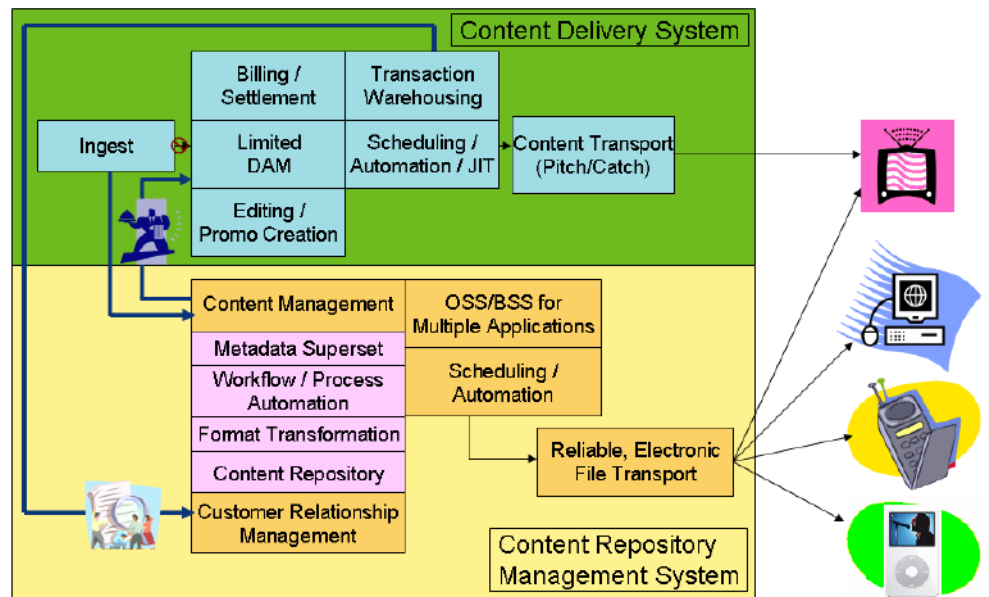
Context—what the content represents—is one essential element for defining the consumer's content preference. (The other is time.) Context may, for example, be the original audience demographic or psychographic that the producer had sought. Most content creators know their target audience. They play the story to that audience to help ensure success at the box office, in the ratings, or at the video store. That target-audience information can easily become part of the metadata describing the content.

The richness of the metadata is critical to measuring habits accurately over time. The movies *Alien* and *The Black Hole* might both be “science fiction,” but drill into the demographics and psychographics of both films and you find that one was created as a mature thriller, the other as a family movie appealing to a younger audience.

In today's VOD environments, most supported metadata focuses only on information needed to present the content to the consumer (genre, director, actors, plot summary, parental-guideline rating, and so forth). Many of these systems do not allow for expansion of the content's metadata set because that act would require a much more robust content management solution. Regardless of what many VOD solutions vendors tell their customers, content management is not their expertise. It's simply a requirement of their service solutions.

Another problem in identifying a consumer's content preference is the need to identify trends in behavior. For example, a consumer may have watched mostly family films all week, but on Friday night, the transaction-tracked choices become filled with murder mysteries, “slasher films,” and modern adult comedies. Does that mean that the next day, Saturday, the content provider should promote *Borat* as a customized offering? Maybe the family's Friday-night babysitter might be interested, but the family who owns the TV wouldn't be. Data from sophisticated customer relationship management and content-trending tools can help avoid such a situation by spotting trends over time that help the content provider ignore anomalies.

A true rich-content management system like EMC® **Documentum**® manages a superset of metadata about the content. It can cover any type of descriptor needed to provide the necessary context. This superset resides outside of the delivery system and can be accessed by all applications that have a need for specific content information.



A Content Repository Management System

For example, today’s VOD systems ingest a package of data that includes actual content, associated “trick files” to simulate fast-forward and fast-review, poster art, and metadata to allow the presentation layer to display information about the content. Most of the time, there is no need for the system to store advanced demographic or psychographic data about the content: The system has no need for it.

If, instead of ingesting content first into the delivery system, the content is initially ingested into a managed content repository, a far richer set of metadata could be entered. This approach allows other open toolsets to access the information. At the same time, a subset of the metadata could transfer to the VOD server in the standard package with all of the other necessary content elements. Such a system would allow content providers to pair the consumer’s transactions to the contextual data in the content repository.

After a consumer’s transactions are analyzed for trends, the CRM tools could again access the content repository metadata to build a unique, personalized offering of content that has a good statistical chance of appealing to that consumer.

Added benefits of a content repository managed by EMC Documentum

When the consumer's transactional and contextual trending analysis is complete, this preference profile could be used to target advertising spots and other messages to the home, increasing the value of the ad spot availability to the advertiser.

Having all content under the control of Documentum brings benefits beyond defining personalized content: Content ingested once into a Documentum-managed repository allows the software to automatically process that content for multiple delivery applications. If content were ingested with a flag giving the distributor the right to present the content via TV, personal computer, or mobile phone, Documentum could automatically kick-off a process flow to transcode the content into the necessary formats, either through software- or hardware-based encoding. Documentum will also create subsets of the metadata for each delivery system, formatted as required by the solution manufacturer.

Documentum also handles the issue of limited-duration broadcast rights. Typically, a content provider will license their content to a distributor for a specific length of time, such as a 90-day window beginning on a certain date. When the contract expires, the broadcaster must delete content from play-out or streaming delivery systems. Documentum's automated workflow engine not only sends commands to perform the delete task, but also automatically notifies the broadcaster's and content provider's legal departments that the content has been purged and the terms of the contract fulfilled.

No database, no service

Personalization of content will become a differentiator among the increasing number of content providers available to each consumer. The company that accurately predicts viewers' tastes, then delivers the appropriate content with good quality of service and cost, will have a clear advantage over competitors. Properly diagnosing a viewer's preferences means providers must track that consumer's habits and choices over a long period. Change providers, and you start the process all over again.

Obviously, databases that track user viewing profiles, choice transactions, and service subscription data are critical to content personalization and user access. These databases will be instrumental to content personalization and will have to be protected, available, fast, scalable, and federated.

In one implementation of IPTV being marketed now, a single database handles service access for up to 100,000 subscribers. If that database becomes unusable for any reason, up to 100,000 customers will lose access to their TV-based content until the problem is fixed. Service-access databases are crucial to revenue streams today, and consumer-profile databases will be nearly as important in the future.

You need to look no further for proof of this than to examine the multi-billion-dollar annual ad revenues that search giant Google generates by providing targeted ads based upon search results. The goal of any content personalization system, therefore, will be to provide targeted content choices to consumers (to keep their attention and subscriptions) while providing a known, targeted audience for their advertisers.

Being able to match the right ad to the right consumer greatly increases the chances that the product being marketed will be of interest to and, hopefully, be purchased by the viewer. It's the difference between a "sniper shot" of information and an un-aimed shotgun blast. Done right, such personalization would deliver an ad for a minivan instead of a sports car to a viewer who recently began watching shows that rate high with expectant mothers and young families.

Being able to target ads appropriately also means that a content provider could charge higher fees to advertisers who want to achieve far better results with targeted ads than they do with traditional untargeted advertising methods. More likely, it will allow the content provider to simply compete for advertising dollars in a world where targeted advertising opportunities are gaining in number. (Remember Google.)

To stay competitive among other providers who also target their advertising, you REALLY want to protect that personalization-data investment—protect it from data corruption, loss, unavailability, or compromise, whether from deliberate acts or system failures. That’s where EMC comes in. EMC provides a suite of software solutions that can help ensure database reliability:

- Multi-site database replication with EMC products like **MirrorView™**, **SAN Copy™**, **SRDF®**, **RepliStor®**, and **Replication Manager**.
- **Double Checksum for Oracle** software to intercept corrupt Oracle data before it is stored in the database.
- **SnapView™**, **TimeFinder®**, **Replication Manager**, and **RecoverPoint** software to provide point-in-time recovery and continuous data availability.
- **EMC’s RSA® Security Division** services to provide data encryption technologies and enterprise-wide heterogeneous security policy auditing and identity validation.
- **EMC Smarts®** software to provide network and application problem identification and root-cause correlation.

However, having a functional database means more than being able to run a simple query. For the system to function well and provide a great consumer experience, the provider must also deal with the response times and scalability of personalization solutions.

As any good database administrator will tell you, the keys to fast database response are:

- Appropriate hardware and software. You can’t run Oracle RAC on a laptop, no matter how much your accountant would like it.
- Queries that are well-written to minimize searching and provide relevant results.
- A reasonable amount of search data—as small a sample as possible that still meets the objective.

The problem with databases, which, by their nature, encompass huge amounts of information compiled over years of use, is that they tend to grow out of hand. Until recently, you had to have all data for a database located in the database itself. But too much data slows down response time.

Obviously we have a problem. On one hand, we want access to as much information as possible about the consumer. On the other hand, we want to minimize that data when doing analysis queries. The solution is to automatically archive out of the database any data not relevant (too old) or not accessed within a set time (such as from a cancelled account). But to most people, archived data means unavailable data. What if the cancelled account reactivates? How do we get that data back into the system?

“Archived” does not necessarily mean “unavailable,” however. **EMC DatabaseXtender®** software locates irrelevant or stale data within databases and moves it automatically to lower-cost second-tier storage. A pointer remains in the database so that a query can still find the data. It’s not gone; it is just not clogging up the main database and slowing things down.

The goal is to be able to satisfy 90 percent or more of all queries with active data, and the rest with a combination of active and archived data. This architecture does two things. First, it significantly lowers database response times. Second, it lowers total cost of ownership by moving rarely used data to a more cost-effective, albeit lower-performing, storage medium. EMC DatabaseXtender software provides the answer to the problem of performing just-in-time personalization analytics on your databases.

The need for content context is universal

Personalization is just one change coming in the way we all will consume content. As we begin to interact with these new content systems and the increased choices they bring, we will also change how we interact with each other.

Just as their parents once discussed top TV shows over the water cooler the day after those shows aired, a new generation of viewers will be recommending new and interesting content instantaneously. Socialization will augment personalization. When the guys tell you how great last night's football game was, you no longer have to regret missing it. It will be available to you on the Web or will be repeated on TV. Or, if someone tells you how good a new TV show is—even when you thought it was something you wouldn't be interested in—you can still sample it to see if you were wrong.

When consumers experience content recommended to them, they may ask, "Is there anything else like this available to me?" Content providers will need to offer search capabilities if they want to stay relevant to consumers. Content context will drive these new content searches and remain the backbone of personalization engines. Today, technology isn't so much a leap forward; it is another step up from the original building blocks already in place. The important step to take now requires building a content infrastructure that can handle change as quickly as it happens.

As new content-socialization techniques emerge (consumer ratings, consumer reviews, consumer-created parodies, fan sites, and more), providers will still need to describe, classify, and package the content they provide to the consumer.

That is content context. And its importance is just beginning to grow.



EMC Corporation
Hopkinton
Massachusetts
01748-9103
1-508-435-1000
In North America 1-866-464-7381

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC², EMC, DatabaseXtender, Documentum, RepliStor, Smarts, SRDF, TimeFinder, and where information lives are registered trademarks and MirrorView, SAN Copy, and SnapView are trademarks of EMC Corporation. RSA is a registered trademark of RSA Security Inc. All other trademarks used herein are the property of their respective owners.

© Copyright 2007 EMC Corporation. All rights reserved.

EMC Perspective
H2689