Designing and Developing Analytics-Based Data Products

Thomas H. Davenport   Stephan Kudyba

September 13, 2016

Abstract

An increasing number of companies are creating products that combine data with analytical capabilities. Creating an effective development process for these data products requires following well-established steps - and adding a few new ones, too.

Thanks to several waves of innovation in recent decades, the rise of information and technology is one of the dominant features of the current economy. The expansion of the information economy that began in the mid-1990s included enhanced hardware and software capabilities, abundant broadband Internet access, and increasingly widespread use of the Internet. These developments helped spur the creation of new products and industries and drove a significant increase in data resources. In an important 1996 article published in MIT SMR, "The Design and Development of Information Products," authors Marc H. Meyer and Michael H. Zack previewed the impact of these changes.[1]

The past 20 years have brought several reconfigurations of the information and knowledge economy, as enhancements in computer processing and storage capabilities, new software and communication technologies, and the evolution of wireless broadband and mobile computing have taken hold. Technological breakthroughs have driven exponential growth in e-commerce and the emergence of a digital economy with vast data assets. The changes have been accompanied by ongoing attempts to make sense of all the data through the use of analytics.

Contents

1 Revisiting "The Design and Development of Information Products"

In a 1996 article in MIT SMR titled "The Design and Development of Information Products," authors Marc H. Meyer and Michael H. Zack examined the design and implementation of information products.[1] Despite the economic importance of the information products industry and the rapid pace of innovation at that time, previous research examining the design and development of information products, which Meyer and Zack defined broadly to include electronic or printed products based on data, information, and knowledge, did not exist.

Meyer and Zack's research was guided by several basic questions: What could companies in information products industries learn from research on physical products? How were information products designed and manufactured, and how can information technology be used to support these processes? And more fundamentally, what was the architecture of an information product, and what were the strategic, organizational, and technical implications of the architecture for companies operating in this arena? To answer these questions, Meyer and Zack built on research and knowledge about the design of physical products. They developed a five-step process for developing information products - acquisition, refinement, storage/retrieval, distribution, and presentation - and showed how companies could create competitive advantage with an information product platform and generate value by refining information.

Obviously, there have been tremendous changes in the information and technology areas over the past two decades - more powerful hardware and software, an explosion of bandwidth, a proliferation of mobile devices, and increasing access to the Internet. But it's instructive to recall where this discussion began and how it has evolved. Today, we argue that it's useful to think in terms of "analytics-based data products" as opposed to "information products." Given the advances in digital technology and the increases in usage, we propose augmenting the Meyer and Zack framework with two additional steps - one focused at the beginning on creating the conceptual product and the other focused late in the cycle on collecting market feedback.

2 About the Research

This research involved two primary efforts. First was a set of 35 interviews with data scientists, many of whom were involved in data product creation, in 2012. The data scientists interviewed primarily worked for startup companies in online businesses, although some data scientists from larger companies (LinkedIn, Facebook, and eBay) were also included. In September 2014, we held a day-long meeting of large, established companies that were exploring data and analytics-based products and services. The participants in this session included executives from State Street Corp., GE, Monsanto, the World Bank, Thomson Reuters, Boston Financial Data Services, EMC, Intuit, MassMutual, Eli Lilly,

Caterpillar, and others. We conducted interviews at many of the companies separately from the meeting, and we also learned about them during the session. We conducted several additional interviews specifically for this article. In total, we interviewed managers at more than 40 companies that had some data product development activities underway. We discussed the types of products the companies were attempting to develop, the processes they employed for development, the technologies and platforms they used to develop and deliver the products, and the success of the products they had in the marketplace thus far.

3   Introduction

The original goal of computerized information was to facilitate internal business transactions and improve decisions. However, the resulting digital assets are now of considerable value in themselves.[2] According to a 2015 Organization for Economic Co-operation and Development research report, "Data-driven innovation forms a key pillar in 21st-century sources of growth ... large data sets are becoming a core asset in the economy, fostering new industries, processes, and products and creating significant competitive advantages."[3]

When digital assets are made available to customers, they are increasingly accompanied by analytics that provide insights and facilitate decisions. Analytics - which were not typically present in earlier information offerings - add substantial value to intangible assets by making them easier to understand and apply. In a world in which information alone has become ubiquitous and somewhat commoditized, analytics provide a means of making information more useful and valuable.

In this article, we will focus on the combination of new analytical capabilities and burgeoning data assets that together form value-added information product offerings. In common parlance, these offerings are often called "data products."[4] For our research, we interviewed more than 40 companies that offer data products and specifically addressed their process (or the lack thereof) for developing them. (See "About the Research.") In our work on this project, we seek to augment and update Meyer and Zack's findings from 1996.

4   The Value of Analytics For Data Products

Data and analytics were historically employed for one purpose: improvement of internal decision making. Indeed, one of the earliest names for these fields was "decision support." The goal was to improve the accuracy and efficacy of decisions in marketing, finance, human resources, and so forth.

But the big data revolution offered companies another use for data and analytics. Beginning with online companies such as Google Inc. and Facebook Inc., companies began to develop "data products" for customers based on data and analytics. Search services from Yahoo! Inc., Google, and others were arguably the first of these, but other sorts of products followed. LinkedIn Corp., for ex-

ample, developed "People You May Know," and soon added others, including "Jobs You May Be Interested In" and "Groups You May Like."[5] Online companies such as Facebook have done similarly. (Facebook has its own "People You May Know" product.) Many online data products are available as mobile apps, shaping what has been called an "app economy."[6] Not every app involves analytics, but those that do qualify as data products.

Data products (most of which can be described as services) are not generally sold separately to customers but are used to attract customers for advertising, draw attention to unknown products in large product pools, and enhance revenue through cross-selling and upselling. They have powered rapid growth in the value and success of these online companies - not only in creating a much larger user base but also in differentiating offerings.

Analytics used in data products take a variety of forms. The most common is descriptive analytics that provide insight on a customer's level of activity or product usage.[7] Google pioneered this format with its Google Analytics offering, a free set of data products that informs customers about visits to their websites. Descriptive analytics can also be comparative, for example, relating a household's utility usage and expenditure to other household's, or comparing a company's travel activity with that of its peers.

Predictive analytics are more difficult to generate. One of the most common forms today is predictive maintenance for industrial machines. Using data gathered from sensors in machines, analytics compute the point when comparable machines have broken down and recommend particular services before that time. Large companies such as GE, Siemens, and NCR have recently introduced predictive maintenance offerings. In the consumer sector, Seattle, Washington-based Zillow Group Inc.'s Zestimate, based on publicly accessible housing data, predicts the price a homeowner might receive for the sale of his or her house.

Prescriptive analytics recommend specific actions. In the agricultural products industry, companies such as Monsanto and DuPont offer data products that recommend, for example, when and what farmers should plant, when certain interventions, such as water or pesticide applications, are advisable, or when to harvest.[8] These data products can help farmers achieve higher crop yields. Other forms of prescriptive analytics in data products involve matching algorithms, which match customers with products, dating candidates, or potential business network members.

Data Analytics:
Descriptive
Predictive
Prescriptive

5   Current Processes and Issues For Information Products

Relatively little is written or known about developing new generations of data products. Indeed, we have noticed that the companies interested in producing new offerings often lack both structure and process. In their 1996 article, Meyer and Zack offered a clear structure for information product development. Although they were primarily addressing traditional information services com-

panies, we believe that many of their ideas are quite relevant to the issue of designing data products from big data, including the following:

- **Product and process "platforms" can support a variety of specific information products.** That idea has become much more popular in recent years, with many economists and strategists extolling the virtues of "multisided" platforms that allow interaction with multiple constituencies at once.[9]

- **Different information products can share an overall product architecture or be part of the same product family.** In Meyer and Zack's view, it's a mistake for organizations to think in terms of single products.

- **An idea "refinery" can create value from information.** Although Meyer and Zack presented this concept 20 years ago, few organizations have begun to extract the kind of value from their information that the authors envisioned.[10]

In this article, we will describe what leading companies are doing to create, refine, and generate value from data products. Given the emphasis that has been placed on the potential value that data assets offer as the digital economy progresses, a revised model for providing structure to the product creation process is in order. Meyer and Zack addressed the topic in the form of an information product development process, where raw data sources or repositories provide inputs to the process of producing a product. They created a five-step methodology (acquisition, refinement, storage/retrieval, distribution, and presentation) to turn data inputs into information products.[11] While we think that model continues to be valid, the evolution of various technologies and corresponding data sources in today's digital economy suggests a need for adjustments and augmentations to accommodate increased volumes, velocities, and varieties of data, as well as the corresponding storage and accessing activities required to manage the assets in the product creation process.

6   An Updated Information Product Model

Data product development activities today are rarely undertaken in a traditional product development sequence that involves identifying the need, developing the product, and then taking it to market. The current pace of business - particularly online business - is too fast for that. Rather, product development activities often take place in a continuous, iterative fashion, with the important activities conducted in parallel.[12] The popular "lean startup" model specifies the development of a "minimum viable product," with periodic refinements over time.[13] To the degree that data product developers are comfortable with any approach to product development, it is that one.

faster time to market

An updated model needs to reflect new "time to market" expectations in the product development process. In light of these new realities, we have added

additional steps to Meyer and Zack's original five-step model. At the front end, we add a step that involves conceptualizing the product; at the back end, we add the establishment of a market feedback mechanism. Additionally, we point to the ongoing need for managers to collect input from a variety of stakeholders.

6.1   Step 1: Conceptualizing the Product

Before jumping in, organizations must identify an information product that meets a need from the marketplace. This introductory step needs to take place before data acquisition. It requires conceptualizing the information product, along with identifying the required data resources. The process involves product definition, data investigation (which should include sourcing data creatively[14]), and establishing the framework necessary to produce a prototype. Once this set of requirements is met, the remaining steps of the development process can be carried out more efficiently. For example, once managers know the data elements that will go into the product, storage and retrieval can be streamlined.

product definition, sourcing data, data investigation, framework to produce a prototype

An interesting example is CarMD.com Corp., an Irvine, California-based company that provides services that leverage automotive diagnostic information. The original idea was to provide diagnostic capabilities that led consumers to auto repair estimates and potential service providers.[15] One of the company's products compares the data extracted from onboard computers in cars against online auto repair databases and offers consumers information on auto maintenance.

6.2   Step 2: Data Acquisition

Once the conceptual model has been worked out, data acquisition can be pursued in a more efficient manner. Organizations tend to acquire or accumulate data that corresponds to their functional activities. However, given the vast amounts of data being generated by information devices and the data available from public sources, the acquisition process needs to connect the requirements of the conceptual model to data that will create the product. In addition to acquiring structured data (for example, customer purchase records), companies should also consider using unstructured sources (for instance, social media comments) that might be able to add value. Companies should be prepared to look within and outside their own systems for such data.[16]

connect requirements to data

combine internal and external data

6.3   Step 3: Refinement

Although Meyer and Zack's data refinement process remains quite relevant, it has to be augmented to facilitate new data sources and to take advantage of advanced analytic methods. The original model talked about the importance of being able to "glean further meaning from combinations of individual [data] elements."[17] Today, much data refining is achieved with automated tools. Real-time machine learning and algorithmic processing of data elements

automated data refinement process: categorize, correlate, personalize, profile, search

7

can categorize, correlate, personalize, profile, and search data quickly to create meaningful models that have significant value for consumers.[18]

For example, Passur Aerospace Inc., based in Stamford, Connecticut, uses both its own data and public data to develop scheduling information for airlines and travelers. Drawing on publicly available data on weather, flight schedules, and other factors, along with its own internal data based on radar statistic feeds, it generates flight arrival estimates. Applying advanced analytics, Passur's arrival estimates outperform ones based on traditional techniques.[19]

6.4   Step 4: Storage and Retrieval

Storage and retrieval are as important as ever. However, retrieval in today's environment must incorporate advancements in query and search processing capabilities (for instance, making use of algorithms) that can access more granular levels of data. Traditional storage techniques need to be augmented by new technologies such as map reduction (a software framework for distributed processing of large data sets on computer clusters of commodity hardware) and parallel processing capabilities to manage larger and faster-moving data sources. Many organizations store data in relatively unstructured formats when they initially capture it, refining it over time. Data storage, retrieval, and processing are increasingly taking place in the cloud rather than on a company's premises. This not only provides companies with flexibility in their technology infrastructures but also can make it easier for them to combine internal and external data.

relatively unstructured formats when data is initially captured, refine it over time

6.5   Step 5: Distribution

The distribution options for information products have shifted dramatically from the earlier menu of possibilities, some of which (such as fax and CD-ROM) have been superseded by the Web. Timing and frequency remain critical aspects of distribution; data products must be continuously available and updated in near-real time. In the digital economy, online media (such as websites and portals) fully address the required level of continuous accessibility to information products. However, Web access via traditional computers is quickly being overtaken by mobile access via smartphones, tablets, and apps. As a result, providers of information products that are distributed via mobile devices need to revamp their content formats and design.

data products must be continuously available and updated in near-real time

At the same time, distributing data products through the cloud adds a new dimension to the question of how frequently information needs to be updated for users. Consider, for example, a business-to-business case involving a shipping service provider that offers information products including en-route metrics, such as estimates of time to delivery. Assuming the data is available, the frequency and timeliness of such information - generated through GPS traffic information, location data, and analytics - can be close to real time.

6.6   Step 6: Presentation

In the original Meyer-Zack model, information products gained value from the context of their use. The user interface mattered - and the easier products were to use, the more valuable they were. Although the digital economy places heavier emphasis on analytics than on simple data provision, there are some important constants. While standard reporting (that is, simple information products) continues to meet the needs of many consumers, more advanced analytics-based products such as forecasts, predictions, and probabilities (such as real-time calculations generated through machine learning) can lead to differentiation and competitive advantage.

6.7   Step 7: Market Feedback

The competitive nature of the information product space, availability of new data sources, and demand for timely decision support require an ongoing emphasis on innovation and on monitoring product usage. Adding this step at this stage of the analytics-based data product development process is consistent with the iterative nature of product development in a "lean startup" context. Once again, the evolution of new technologies has provided a mechanism for facilitating a feedback and information extraction process from the marketplace. New forms of market research are capable of leveraging social media platforms (for example, business Facebook pages) to listen to the marketplace. Interactive blogs and flash surveys can be utilized to assess customer perceptions of existing information products.[20] New features of online information products can be tested in a matter of hours with A/B or multivariate online testing approaches. Both user correspondence and digital metrics on product use (for instance, views, clicks, downloads, and bounces) can be analyzed to enhance products continuously.

7   A Structured Approach to Stakeholder Involvement

In order to achieve effective results from the implementation of the product development model, stakeholder involvement is essential.[21] Having particular types of input at different stages of the product development process is important. Therefore, companies need to develop some degree of structure for stakeholder input.

During the stage when the product is being conceptualized, it's important to have involvement from three specific groups: subject matter experts at the business level (who can help determine the feasibility of the product design); managers of existing and complementary information products (who can help companies avoid cannibalization and duplication); and marketing people (who can help assess the nature and scale of consumer demand). These individuals can

assist in providing the framework for designing or upgrading existing products to add value to meet market needs.

For the data acquisition and the storage and refinement stages, stakeholder involvement should expand to include legal representatives, who can speak to data ownership, privacy, and use issues; IT personnel, who can provide input on hardware and software requirements for data products and also help in developing and improving the functionality of the product; and data managers and analytics and data scientists to assist in product platform execution. It is critical to involve analytics and data science professionals to help in structuring and analyzing data.

For the distribution and presentation stages, the stakeholders should again include marketing people (who can help sort out consumer/user needs for the initial product launch and subsequent product releases) and IT personnel (who can deal with hardware and software issues in product functionality during the product rollout).

During the market feedback stage, it's important to involve people from both IT and marketing. IT personnel will be able to leverage available communication devices to interact with users on a variety of platforms (Web, mobile, social media, email, etc.). Marketing can devise strategies for interaction and feedback and will be able to gather and synthesize the feedback.

If data product development is to be successful, human resources personnel also have a pivotal role to play. Although data scientists can be difficult to hire and retain, they are essential for developing analytics-based data products.

8   Structure Versus Market Responsiveness

There has always been a trade-off in product development between having structures that ensure that the product addresses market needs and is of high quality, and being able to introduce products quickly and remain responsive to customer needs. Though the pendulum in data products has clearly swung in the direction of responsiveness, there is still a need for structure and method in developing new offerings.

As the proliferation of information has led to commoditization, it is often the accompanying analytics that make data products truly useful. Analytics can be difficult and time-consuming to develop, so it is vital to have a sense of which analytics are needed and valuable before developing and introducing them. We believe that the steps outlined in the Meyer and Zack article - combined with the additional steps we outline - can be helpful to organizations in assembling data and analytics products that provide value to consumers and businesses.

Just as in the early e-commerce era almost every company decided it needed a website, we envision an era in which almost every company feels the need for a data product. It will thus become increasingly important for organizations to have some discipline about which data products are developed and which functions they incorporate.

We doubt that Meyer and Zack could have foreseen the explosion of information and technology that has taken place since 1996. However, their article provided early insights and warnings about the importance of rigorous thinking when developing products based on intangible information assets.

9   ABOUT THE AUTHORS

Thomas H. Davenport is the President's Distinguished Professor of Information Technology and Management at Babson College in Wellesley, Massachusetts, and a fellow at the MIT Sloan Initiative on the Digital Economy. Stephan Kudyba is an associate professor of data analytics and management information systems at the New Jersey Institute of Technology's Martin Tuchman School of Management in Newark, New Jersey.

10   REFERENCES

1. M.H. Meyer and M.H. Zack, "The Design and Development of Information Products," Sloan Management Review 37, no. 3 (spring 1996): 43-59.

2. P. Rotella, "Is Data the New Oil?," April 2, 2012, www.forbes.com.

3. Organisation for Economic Co-operation and Development, "Data-Driven Innovation: Big Data For Growth and Well-Being" (Paris: OECD Publishing, 2015).

4. J. Howard, M. Zwemer, and M. Loukides, "Designing Great Data Products" (Sebastopol, California: O'Reilly, 2012).

5. T.H. Davenport, "Analytics 3.0," Harvard Business Review 91, no. 12 (December 2013): 64-72.

6. Organisation for Economic Co-operation and Development, "The App Economy," OECD Digital Economy Papers, no. 230 (Paris: OECD Publishing, 2013).

7. The three types of analytics are described in detail in T.H. Davenport and J.G. Harris, "Competing On Analytics: The New Science of Winning" (Boston: Harvard Business Review Press, 2007).

8. J. Bunge, "Big Data Comes to the Farm, Sowing Mistrust," Wall Street Journal, February 25, 2014.

9. A. Hagiu, "Strategic Decisions For Multisided Platforms," MIT Sloan Management Review 55, no. 2 (winter 2014): 71-80.

10. Meyer and Zack, "Design and Development of Information Products."

11. Ibid.

12. One CEO creating a data product in the health care industry told us, "We tried agile [referring to agile product development methods], but it was too slow."

13. E. Ries, "The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses" (New York: Crown Business, 2011).

14. D. Barton and D. Court, "Making Advanced Analytics Work For You," Harvard Business Review 90, no. 10 (October 2012): 78-83.

15. Bain & Company, "Using Data As a Hidden Asset," August 16, 2010, www.bain.com.

16. E. Dwoskin, "Startup Factual Knows Your Commute, and Much More . . . ," Wall Street Journal, Dec. 10, 2015.

17. Meyer and Zack, "Design and Development of Information Products," 46.

18. D. Kiron, P.K. Prentice, and R.B. Ferguson, "The Analytics Mandate," May 12, 2014, https://sloanreview.mit.edu.

19. A. McAfee and E. Brynjolfsson, "Big Data: The Management Revolution," Harvard Business Review 90, no. 10 (October 2012): 60-68.

20. B. Zeidler,"6 Ways to Extract Customer Insights From Social Conversations," February 2015, www.quirks.com; and T. Rajpathak and A. Narsingpurkar, "Managing Knowledge From Big Data Analytics in Product Development," white paper, Tata Consultancy Services, Mumbai, India, 2013, www.tcs.com.

21. B.H. Wixom, "Cashing In On Your Data," research briefing, MIT Sloan Center for Information Systems Research, Aug. 21, 2014, http://cisr.mit.edu.