

Data as an asset *a systematic appraisal*

Cetin Karakus

Version 1.0, Aug 2020

Is data really an asset?

An asset, using the common definition of the term as derived from the corporate finance theory, is characterized by its ability to generate net positive cash flows and thus has a positive net present value. Everything obvious about the data we know of point to the fact that it is a negative cash flow generating entity: it costs to acquire it, clean it, store it, secure it, back it up, serve it and the list of negative cash flow generating activities associated with managing the data just go on. By account of this, data can best be described as a liability rather than an asset. What is it we are missing here?

The short answer is we are not missing anything and the data is really a liability, but a liability with the potential to be *transformed* into an asset. The size of the potential and the ease with which to realize that potential varies greatly across different kinds of data an organization deals with. Operationally significant data such as customer orders, scheduling and shipping data, sales inquiries, etc. have more "obvious" benefits and such benefits are likely to be the driving force behind the major parts of a business organization's IT systems and applications landscape. Thus, there is a compelling case for the operational data being an asset—in spite of the all the costs involved in managing such data-- because it powers the positive cash flow generating core business activities.

The case for myriad other types of data an organization deals with on a daily basis is a lot less clear and it forces us to think about what is that makes data to have the potential that positions it to be considered as an asset. To that extent, there are two fundamental principles—akin to the fundamental theorems in a mathematical theory—that can underlie our discussion on this matter:

First (fundamental) principle of data analytics

Data is an asset only if it can be turned into analytics. As a straight-forward corollary, if data cannot be turned into analytics—for whatever reasons—then it is not an asset. In fact, there is even a stronger statement due to the fact that there are significant costs associated with managing data: *data that is not an asset is a liability.*

The importance of implications of the above statements cannot be overstated. Once you see the data being valuable only when it can be turned into analytics and become aware of all the "fixed costs" associated with keeping it around, you get into the right frame of mind where you actively seek ways to turn data into analytics. That will lead us to the second and less straight-forward principle.

Second principle of data analytics

The potential value of data can only be discovered through an marketplace mechanism. This follows from the success of free market economics where the prices—and values by proxy—of billions of products and services are reappraised and constantly dynamically adjusted. Contrast this with a command-control type economy where the prices are "pre-ordained" by a central committee. The analogy might seem far fetched but consider how data sourcing is usually done in a big organization: not by the direct involvement of the users of the data who can appraise the data and

assign realistic values to it based on their intimate knowledge of working with it and seeing the potential uses of it. It is usually done by a group in an IT organization and in a way that is based on that group's subjective assessment of what is valuable and what is not.

A direct consequence of the second theorem is the suggestion that there has to be a marketplace where data as an asset can be subject to the relentless forces of supply and demand and its value generating potential can be discovered along the way. I call such a marketplace as a *enterprise data exchange*.

Enterprise Data Exchange

Just like a financial exchange allows buyers and sellers to trade a set of financial instruments with well-defined characteristics through a supply/demand mechanism, an enterprise data exchange is a platform where producers and consumers of data meet to "trade" *datasets* which are clearly identified, structurally well-defined, potentially metadata annotated and hopefully well-documented data artifacts.

The mere concept of the enterprise data exchange as defined above allows a quantum leap when it comes to the enterprise data management practices:

- **Data Identification:** A data exchange forces the data stakeholders to clearly identify the data artifacts that are going to be "listed" on the exchange. As mundane as it may seem, this is a very important first step and introduces a certain type of rigor and discipline in data management processes. Moreover, an exchange is a live platform and the implied dynamism of the platform forces the adoption of organization-wide data identification processes that are directly linked to the primary sources of data and hence can stay up-to-date even when the data at the source changes. Just like no one wants to place orders for de-listed companies in a stock exchange, no user of a data exchange would want to deal with the datasets that got out of sync with their sources and hence diverged from the source data structurally or semantically. Such a strong user preference establishes the required discipline on data providers' part to keep their datasets up-to-date all the time.
- **Data Discovery:** Once data is clearly identified through a robust framework and kept up to date with the dynamic organizational changes on a regular basis, then it becomes straight forward to "discover" new datasets. Just like financial exchanges provide reference data for the instruments they provide on their platforms, an enterprise data exchange will have a dynamic catalog of datasets available through it. This catalog should not be confused with the often-ill-fated enterprise data cataloging efforts seen elsewhere. The data catalog I talk about is a living and breathing entity, placed at the hearth of the data exchange and kept up-to-date by necessity as opposed to "the data-cataloging as an exercise to create a database of the data fields in an organization".
- **Metadata Augmentation:** Just as parsimonious transaction details of a trade executed on an exchange is often "remedied" through trade enrichment processes as a part of post-trade processing, datasets listed on a data exchange can be augmented with various kinds of metadata. Such metadata augmentation can not only help with the primary "pre-trade" activity of data discovery but also the "post-trade" activities such as utilizing it to enhance the underlying dataset structurally and semantically. Having metadata capture facilities in the data exchange allows two major benefits. First, it enables concentrating organizational metadata

capture and enhancement efforts—which are often pursued as separate major activities in their own rights with similar tragic consequences as the data cataloging efforts—in a single place where the value of the metadata can be best utilized thanks to the pre and post trade use cases mentioned earlier. Second, as the metadata is likely to play a key role in transactional (buy/sell, provide/subscribe) and data usage tracking & value appraisal processes for which the data exchange is responsible for, having metadata captured and managed at the exchange level is both a logically sensible choice and also a technically efficient way that avoids the unnecessary integration with the external metadata management systems that would otherwise have been the case.

- **Data Provisioning:** Exchange model of bringing together consumers and providers of data lends itself naturally to a subscription model where consumers of the data subscribe to the datasets they express interest in. This subscription model of connecting users of data to the providers do not presume any prior financial arrangement such as having to pay for the subscription fees etc. It may well be the case that the entire data operations are centrally funded across an organization but you may still want to have a subscription model where you can both have a control point to decide whether a specific demand for data use is acceptable—considering various aspects of compliance, privacy, commercial sensitivity, etc.—and also be able to maintain a long-term control through mechanisms such as time-limiting subscription periods and periodically reviewing them.
- **Data Usage Tracking:** Once a subscription model as described above is in place to provision the datasets listed on the exchange, it becomes important to track the actual consumptions of the data. Before moving on to how to do that, note the following benefits data usage tracking provides. First, it will give you the vital statistics on what is most popular and cared about by the people in your organization. This could be a good proxy for *high-value data* you should pay special attention to. Second, it can allow you to discover your organization's *data map*: which datasets being used where—systems, applications --, by whom—users, groups, teams—and when—times of days, days of week, occasions of business event calendars. Third, you can do trend analysis over the usage patterns and examine the dynamic evolution of the organizational data map over time and may be able to draw interesting conclusions about the specific significance of certain datasets in the overall organizational strategy and the associated changes such a strategy might have been driving. This will be an exercise in *data-as-strategy* par-excellence. Finally, the ability to collect vital stats on the what/where/when/whom aspects of data usage on the exchange platform can form the basis for "let-the-best-one-win" data competition strategy whereby instead of wasting valuable resources on deciding which team is best suited to provide the important datasets the organization needs—i.e. picking the winners in a command-control style economy--, you let "the market" (i.e. the data consumers in your organization) to decide with the obvious benefits such an approach entails.

On the question of how data usage tracking can be done, there are good and bad news. The good news is it can be done in such a way that it can enable the value realization agenda mentioned above. The bad news is that it is neither easy nor completely transparent—as far as the potential changes required in data usage sites are concerned. The full technical details of how it can be done is beyond the scope of this paper but I will mention one part the of the solution that can give you some ideas about what is involved in the overall effort.

Any data usage tracking needs to be tightly integrated with the data provisioning and subscription technology. One can use a software licensing style subscription model whereby data consumers are issued *subscription licenses* upon being provisioned with the datasets they requested. If we ensure any data delivery is done through a service-based architecture whereby

datasets are served exclusively through the data service APIs — this might not be the case for every organization and may require a major data engineering work to make it happen but any such efforts will be well worth it. This way, not only we can have fine control over the access to the data but we can also perform all of the data tracking objectives mentioned earlier by simply requiring the subscription license keys --which can tell us about the who/where/what aspects of the consumers of the data-- being part of the data service requests.

Data Monetization

Another great benefit of having a data exchange platform is how smoothly and naturally it can lead to the realization of the hottest topic of data analytics space: *data monetization*. Data providers can trivially put price tags on the datasets that they make available on the data exchange, and the exchange itself can deal with the internal accounting of who owes whom how much. All this is rather obvious. What is perhaps not so obvious is the following avenues of developments that can form the backbone of an effective monetization strategy:

- **Effective Price Discovery:** Supply/demand mechanism inbuilt in the data exchange allows data providers to experiment with their pricing strategy as opposed to going with their gut feeling and setting the prices arbitrarily. Experimentation with the pricing is key to discover the right price that reflects the value of the data as perceived by the consumers of the said data. This can be done, for instance, through setting different prices along the demand elasticity curve and trying to maximize the total revenue by choosing a price point that corresponds to an optimum mix of high price and high demand that translates into the highest revenue.
- **Build/Buy/Sell Analysis:** Cost analysis of the datasets listed on the data exchange coupled with the discovery of the prices they can command make it possible to conduct ROI type benefits/cost analysis. Such analysis is quite valuable both for helping with deciding whether to source (i.e. buy) the datasets rather than build them internally — assuming the building is a possibility, which may not be true most of the times — and also whether to sell the datasets built internally to the external parties outside the organization. Selling internal datasets to the outside parties is obviously a lot more involved process than doing the same with the internal parties, and thus it requires further elaboration on certain key aspects such as looking into the proprietary nature of data, checking the relevant compliance regulations, listening to the privacy concerns, paying attention to the security protocols, ensuring a competitive pricing, etc. Nevertheless, the internal cost analysis and the organization-wide price/demand discovery process are two great tools that are likely to play a key role in the expansion into the outside "markets" should the data providers choose such a strategy.
- **Data Value Chain:** Subscription licensing model that oils the wheels of the data exchange is flexible enough to allow not just the basic business model based on charging the data users for the consumption of data but can also enable a more franchise-type business model — akin to the OEM chains in PC world-- whereby different teams in an organization can build over each others' datasets, resulting in the formation of a *data value chain*. Data subscription licensing and usage tracking arrangements natively supported the data exchange can facilitate this process and deal with the complex mechanics of transfer pricing and profit/cost dispersion along the value chain.

Summary

Data is a liability unless its positive cash flow generating potential is unleashed by means turning it into useful analytics such as the case when the data captured while carrying out core business activities is turned into operational insights that can power the efficient and the profitable operations of a business.

True value of the data can only be discovered through a data exchange platform that brings together the key pieces of the enterprise data management in a single place: data identification, data discovery, metadata augmentation, data provisioning and data usage tracking. Once such an enterprise data exchange platform is in place, the path to the data monetization is wide open, with the further possibilities of achieving effective data pricing strategies with natural build/buy/sell pivoting dynamics and the ability to create data value chains that could go far beyond the basic end user focused subscription model.