DFM DATA CORP.

Interconnected. Interoperable. Invaluable.

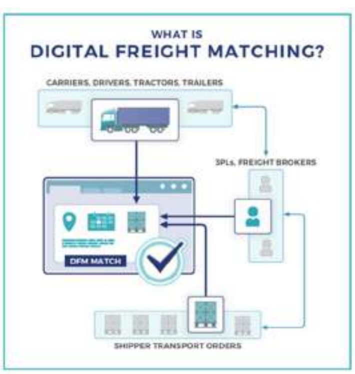
OUR MISSION

"Driving Collaboration Between Digital Freight Matching Providers"

Subject

Dealing with Phantom Load Data in a Dynamic Freight Matching (DFM) Systems. A "**Phantom load™**" can be defined as a specific load record "data record or object" which does not accurately portray a real-life "load of goods". This Phantom load can exist in any logistics system which creates, uses, processes, distributes or matches these loads for the purpose of moving goods from Point to Point. More specifically, a Phantom Load can exist for a multitude of reasons, most proper, some not so much. Phantom Loads can be the result of poor data hygiene, invalid or inadvertent duplicated load data or may be the result of a more menacing purpose. This White Paper will set forth a first or "low hanging fruit" project as a test and kickoff of a consortium focused on the advancement of collaboration and cooperation of logistics entities to work together for the common good and advancement of the Logistics industry. This first project of the DFM Data Corp Consortium is the culmination of many to bring together the Logistics industry in a collaborative fashion to create collective cooperation and competition.

Project Background



WHAT IS DYNAMIC FREIGHT MATCHING:

Dynamic (Digital) Freight Matching is a process or service that saves shippers, carriers and brokers time and money by using technology to dynamically connect shippers to available driver-tractor-trailers that match their specific parameters for load specifications.

For the purposes of this whitepaper, the term DFM will be used to mean any technology, system, service, company or complexity thereof which when exercised together produce an intrinsic load/truck asset match of available resources on both sides available for rate confirmation and load assignment to a specific truck/carrier for fulfillment. The pace of DFM adaption has seen a sharp upturn due in-part to the tragic introduction of Covid-19 into the global ecosystem and our day-to-day life.

To that point, driving without a load, or not being able to find a load is as good as losing money, so keeping freight onboard as often as possible is key to making the most of the drive time for carriers. Dynamic Freight Matching technology gives drivers the opportunity to have load opportunities sent directly to them and gives shippers/brokers with all the automation in between the opportunity to find and match freight and truck anywhere, anytime.

New technologies and efficiencies are moving rapidly. Yet, the adoption of dynamic freight matching services by shippers, carriers and brokers is not growing as quickly as some had anticipated. By mid-2020, some analysts estimated that Dynamic Freight Matching service providers (DFMs) had penetrated the $800B+ truckload (TL) freight market by less than 5%.

FACTORS INFLUENCING MARKET ADOPTION RATE:

Today, the customer experience (CX) of dynamic freight matching platforms and marketplaces communicates an imperfect solution. The imperfect CX combined with generally low awareness of DFM services, and persistent technical barriers associated with a lack of data-sharing, uninformed marketplace (carriers and brokers) and standardization amongst market participants, converge to significantly influence a low rate of adoption of DFM services.

Contributing factors include:

- **Phantom Load and Capacity Data** - Nearly all DFM data platforms are polluted with what appear to be bookable trucks and loads that may have already been booked on another platform or marketplace. Each time this phantom data is perceived as a false positive match, it wastes labor and degrades the customer experience. Worse, phantom data is growing exponentially and inhibiting overall market growth.
- **Siloed Marketplaces** – The DFM industry is a collection of siloed marketplaces that lack the interoperability and data standards to identify and cleanse phantom data across the many DFM platforms. In-fact, many use this confusion to their advantage to move rates per mile whereby increasing profit by means of bad acting.
- **Low Awareness of DFMs** - Hundreds of thousands of carriers and more than a million shippers currently are not leveraging dynamic freight matching services. It’s not that they prefer legacy low-tech best practices, they simply lack an awareness of the opportunity or they may have had a bad experience with a technology provider, with phantom data or both.

A System for Cleaning Phantom Data

System Architecture

DFM Data Corp. has developed a central data clearing house and an anonymized data hygiene service to clean up phantom data problems, industry wide.

In the proposed system, each participating DFM will send JSON/YAML text messages to DFMDC when contracts are booked (accepted rate-con) between loads and trucks. DFMDC will re-broadcast the notices to the other DFMs in the Consortium. Those DFMs can compare their listings against the notifications to see if they match. If a match is evident, each DFM can decide how to use the information in their operations.

DFMDC will supply the software to operate the notification system and it will manage a Data Lake to collect transaction data, both as a backup and as a platform for future additional analyses in the same way other transportation consortia models function (ex: travel – Airline Tariff Publishing Company) ATPco – see document ATPCO for more comparison detail.

Data Hygiene Strategies

DFMDC has looked at the structure of the freight booking ecosystem and the confusion and waste caused by phantom, inaccurate and fake data and DFMDC believes there are several ways to address the problem:

- **Rely on DFM Responsibility**: The current approach is to ask organizations or entities that post loads to take responsibility for clearing up outdated, phantom or false listings that are left behind. To be fair, many participants (e.g., the TMS software providers) make a conscious effort to do this. However, consideration and compliance vary across the industry and some participants cannot deal with this issue head-on. Beyond that, there is a growing segment of the DFM industry that relies on automated re-postings of loads and trucks. They accept listings from Transportation Management Systems (TMSs) and from other DFMs or they push their own listings to other load-boards and independent brokers as a service to customers. There are many reasons for this and a separate conversation as to when and why this happens is detailed - see "Load-board definition" document.

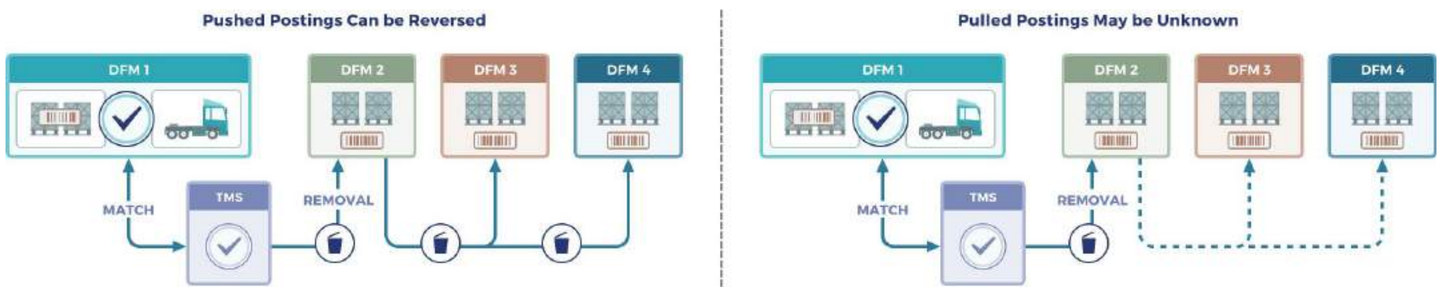


Figure 1 - Even Automated Postings Can Get Lost

As automated "robotic" posting increases, it will become harder for the "original" booking sources to understand the complete visibility of the booking or even know where all of the listings are, in-fact visible, and from which visibility point. Removing the completed, invalid or phantom postings will become harder still, even for those companies that wish to be diligent and responsible.

- **Collect and Analyze DFM Industry Activity**: A direct solution to the problem might involve a central clearinghouse where DFMs mirror all of their listings and bookings (trucks and loads). The central facility could apply matching logic (and "likely" machine learning techniques) to recognize duplicate listings and signal the relevant DFMs with "disposition and hygiene" information. The two drawbacks of this approach are a) it would require member DFMs to place a large degree of trust in an external entity and b) it would likely be costly and time consuming to develop a cleansing mechanism for data from so many sources.
- **Systematic Anonymous Cleaning**: A third approach is to design and install a targeted system to identify and remove obsolete and redundant material with minimal sharing of business transaction details. The key is to engage the DFMs to perform some minimal data preparation and cleansing before submitting transaction information to the system. Theoretically, a combination of a bit of data preparation and distributed comparison and cleaning could remove most of the inaccurate listings. This approach requires industry cooperation and coordination, but member DFMs do not have to trust competitors with their business details. Creating and maintaining a system like this is the responsibility of a neutral third party who is unincumbered – hence DFMDC's founding mission.

To systematically clean the phantom data from the DFM marketplace requires two, linked components:

- A method to compare listings from different marketplaces to determine which listings represent the same real-world loads and trucks. The comparison method should not require the exchange of sensitive business transaction details between and among competing DFMs. – Load/Truck matching logic, aka, **Asset matching logic**
- A system to convey that information to participating Consortium members for action. – **A booking notification system**.

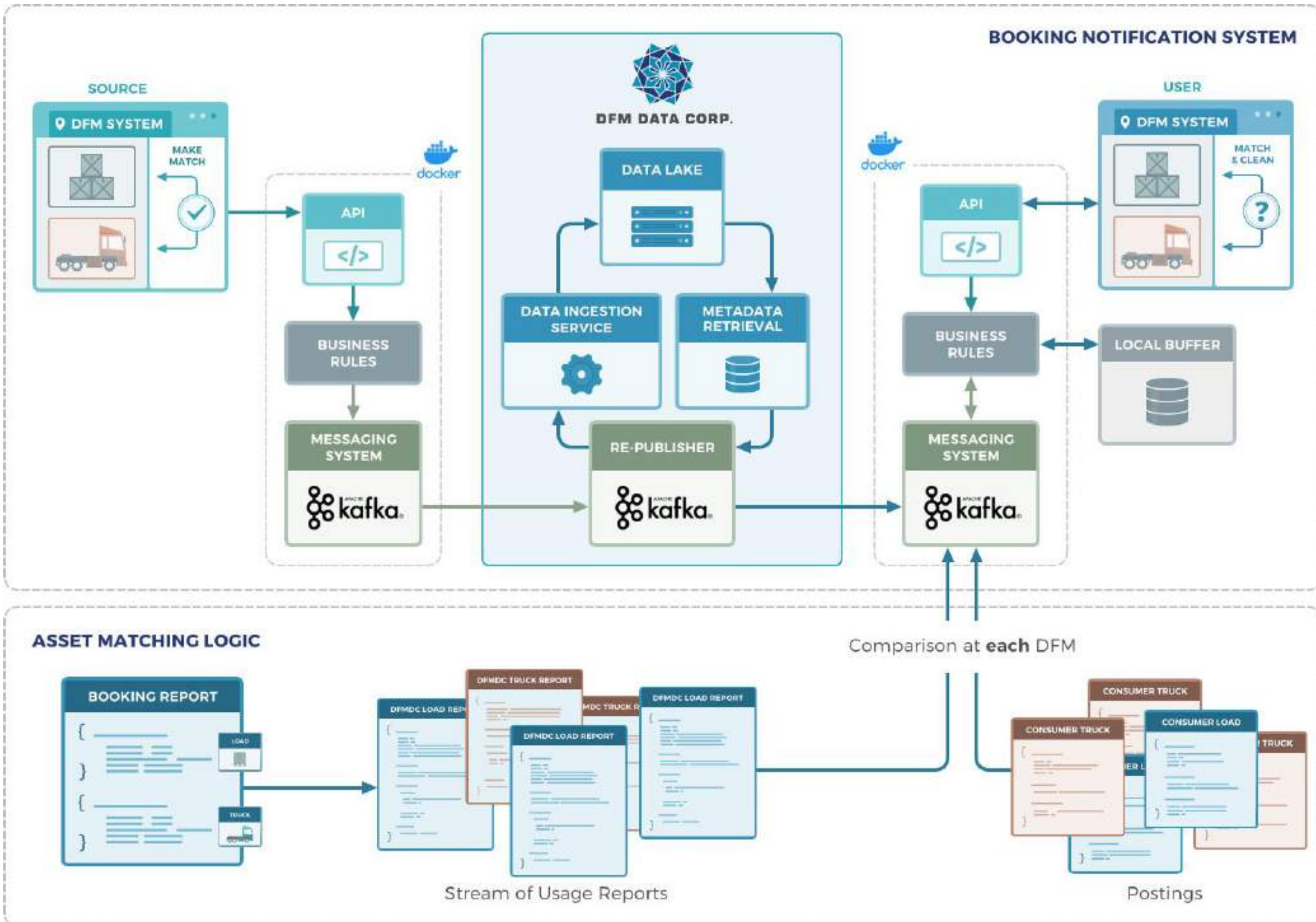


Figure 2 – Garbage Collection System

Anonymous Matching Logic

While member DFMs may want to remove expired, invalid and phantom data from their systems, they are not eager to share their business data with competitors or customers. Any proposed solution must somehow hide the business details. One way to achieve this is to use one-way hashes.

Once a load and a truck are booked on a DFM board or in a marketplace, the challenge is to identify the same load or truck (possibly separately) on other DFMs. We need some common ID, label or the concatenation of multiple pieces of information which can be uniquely recognized and compared.

That unique ID must be created for the specific load or the specific truck/trailer and captured every time the load or truck are listed on a marketplace. Assigning IDs later in the DFM network won't help. The tie back to real-world assets will be missed. Assigning an ID in DFM Data Corp's cleaning system is even less useful.

This poses the first critical system design question: ***Where do we find a unique label for each load or truck that can work across all DFMs?***

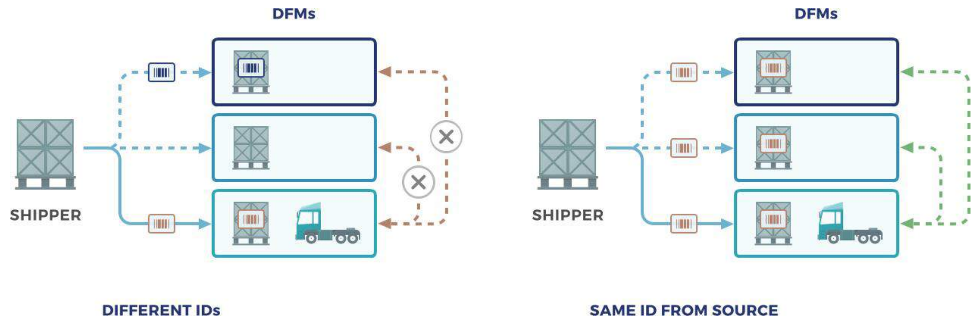


Figure 3 – The Origin of The ID Matters

To date, DFMDC has not identified a universal ID candidate. Instead, there are several possibilities, that are used frequently, but not pervasively. A universal ID may become clear as more examination is done over additional data, or evolve over time as other standardization efforts grow. For the moment, however, DFMDC is looking at a strategy to employ a "collection" of partial solutions. This section outlines some of the high-level options and more detail is provided in Appendix 1.

Intrinsically Hard IDs – a definition:

The trucking industry often deals with assets that have been labelled in a very specific way. Power Units have state issued license plates. Drivers have mobile phones with unique numbers. Carriers have MC and DOT numbers. Shipments may be assigned an ID under the GS1 standard.

If these IDs are formatted in a standard manner and hashed using a standard algorithm, the resulting hash string will be incomprehensible, except when it is compared to another hash string. If they are identical, it is certain that the two assets have the same original hard ID. Only the asset owner will know what the ID is in clear text, but anyone comparing the hashes will know "only" that the two assets (the pair) are the same.

In a perfect world, the entire logistics industry would adopt one standard identification method and DFMDC's system could operate with just two hard IDs, one for the truck and one for the load. In practice, however, the industry hasn't achieved that level of consistency and there are several competing candidates in most logistics situation.

DFMDC's response to this inconsistency is to collect and hash as many hard IDs as the shipping parties will provide. This is possible because matching hashes of any hard ID is conclusive evidence that the items are the same. If their GS1 IDs match, it must be the same physical load. If the mobile phone numbers match, it is the same driver's mobile phone. We don't need to match all of the ID hashes that we collect. Matching any one of them is sufficient if they are indeed hard IDs.

Hardened IDs – a definition:

Unfortunately, many of the common identification systems cannot guarantee enough uniqueness. For example, a shipper may assign a barcode shipping number to each load. If the shipping number comes from their internal software systems, it will be unique for that shipper. However, it is possible that some other shipper may use the same software and at some point, assign the same shipping number. How would DFMDC tell them apart if the hashes matched?

In practice, that is probably not a common occurrence, but with a little extra work we can make it virtually impossible. The trick is to combine the shipper or trucker's ID with other data that describes that particular load. Even if the ID might have duplicates, the combination of variables is nearly certain to be unique. This is illustrated in the following table.

Variables	String to be Hashed	MD5 Hashed String	Comments
SAP order #	"23232323"	9120163167c05aed85f30bf88495bd89	Duplicates are possible
SAP order # + destination zip	"23232323+30303"	59f89bc60f802bdb34edc4b7117a10bd	Duplicates much less likely
SAP order # + destination zip + pickup day	"23232323+30303+011221"	4e8b769bd11e1d69a085ff37919487f9	Duplicates are virtually impossible.
SAP order # , destination zip + pickup day	"23232323,30303+011221"	ce0158f76ee8ff54d94500053d0d2ea4	Changing + to comma totally alters the hash.

Table 1 - Hashing Examples

The challenge with this form of "hardening" is to ensure that all parties construct the input string to the hash in the same way. To succeed, **DFMDC and its Steering Committee will have to design and publish a standard that lists the permitted variables and gives precise instructions on how to construct them**.

Constructing Soft IDs – a definition:

In some (initially many) situations, it may not be possible to obtain hard or hardened IDs from the participating DFMs. Their contributors (manufacturers, shippers, brokers, carriers, truckers and other DFMs) may not have the unique values or may initially be unwilling to share them.

This does not mean that the Matching System is useless. Shippers and truckers must share at least some information simply to consummate a booking and that may be enough to help identify many duplicate postings and loads. The following table of real data shows the results from analyzing 30,000 booked loads listed by two DFMs. Both DFMs listed the origin and destination zip codes and the required pickup date.

Number of Loads that Recur	Percentage of Loads with # of Occurrences
	oZip+dZip
1	59.6%
2	11.3%
3	5.5%
4	4.2%
5	2.7%
6	2.2%
7	1.6%
8	1.5%

Table 2 – A Few Variables Can Distinguish Many Loads

This is a small set of bookings, but it shows that 60% of the booked loads were described by a single combination of origin zip and destination zip. If a third DFM had a load in its marketplace with that specific combination, it is very likely that the load was the same as the one in the booked shipment.

Normally, the third DFM would not know that its load posting with this soft ID data matched a booking in another DFM or on another load-board. However, DFMDC can communicate the soft ID of matched a booked load and, because it is the central clearinghouse, it can give a running count of how often that soft ID has been seen on recent booked loads of the connected systems.

Multi-Element Matching Criteria

The overriding problem with cleaning DFM data is the fact that every DFM, as well as their contributing shippers, brokers, carriers, truckers and TMS users, have their own preferred data schema to describe loads, trucks and booked shipments. As a result, it is currently hard to see a consensus on which hard or even hardened IDs can be used. There is no universal hard ID that everyone uses. At best there may be a set of hard IDs from which most participants might choose to record one or two. The situation with hardened IDs is similar.

When a DFM books a load, it sends (in the case below) snippets of JSON/XML code to the DFMDC with all of the different IDs (hard, hardened and soft) that it has collected concerning the shipment and the truck. Different DFMs may send a different set of ID candidates, and to preserve business privacy, most of the items in the collection are transmitted as one-way hashes (e.g., MD5) for the ID values.

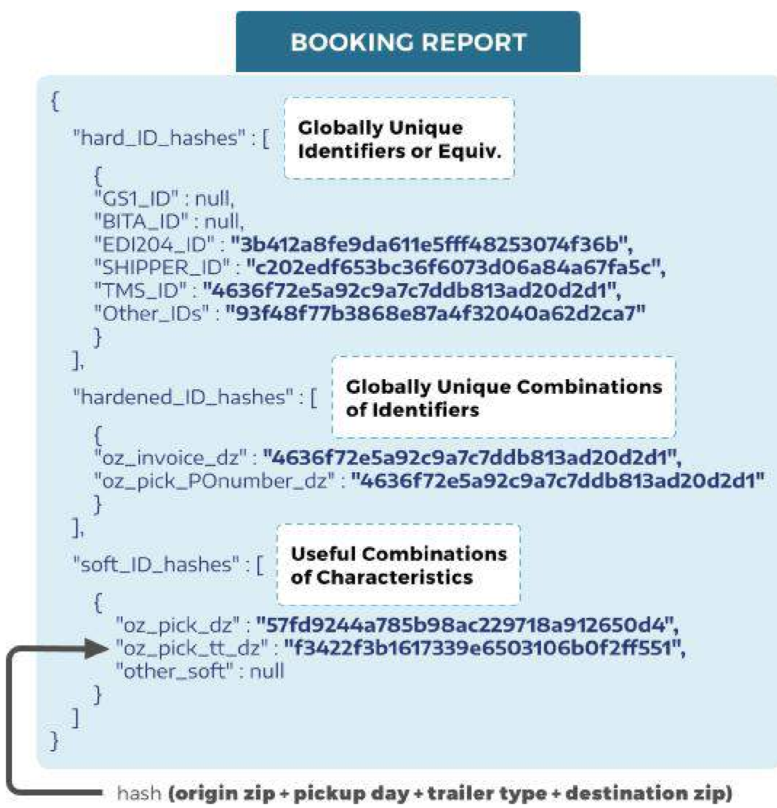


Figure 4 - Multi-Element Booking Report (load part only)

Figure 4 illustrates the general type of JSON/XML structure that might be used. The snippet collects hashes for all of the IDs that happen to be available for a given load or truck. The hope is that other DFMs will collect at least some of the same information about the load.

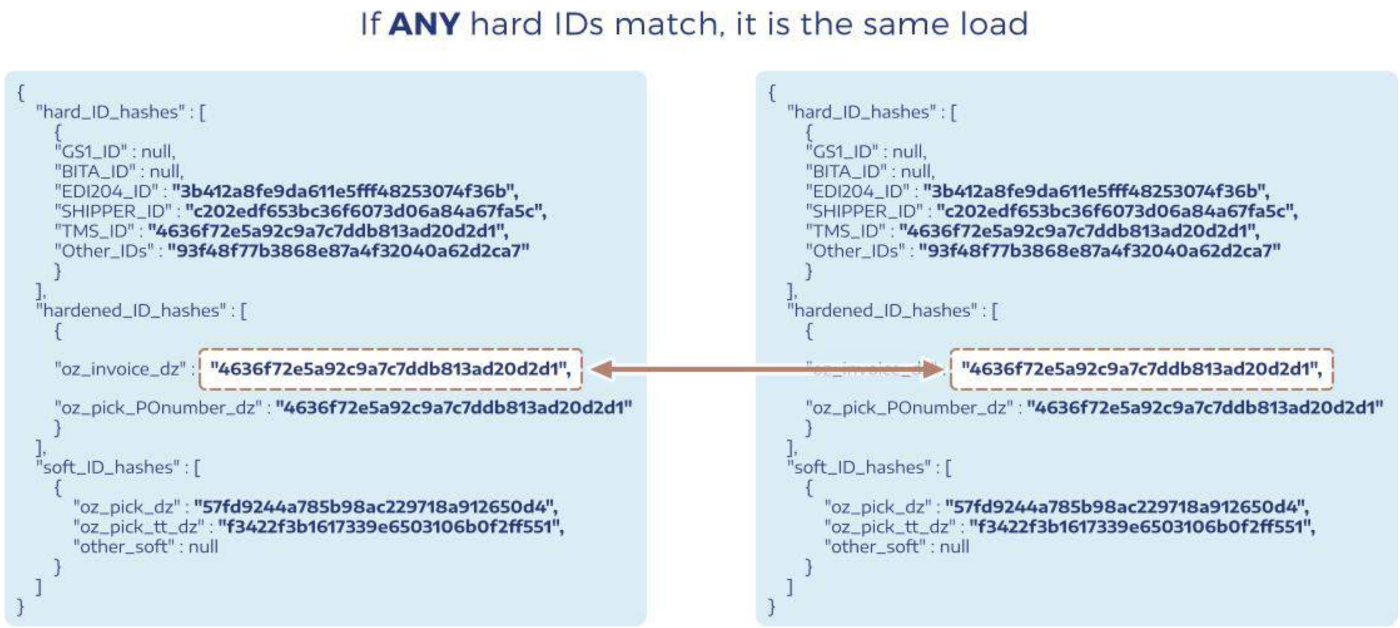
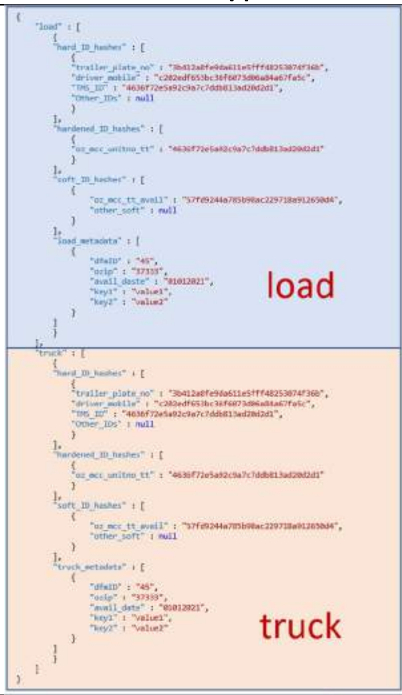
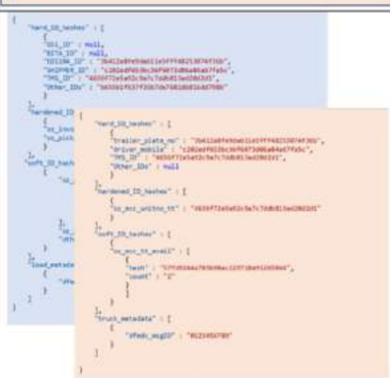


Figure 5 - ANY Hard or Hardened ID is a Definitive Match

The key insight is that if ANY of the hard or hardened IDs match, it is very likely the same physical load or truck, **so the matching strategy is to collect everything, including the kitchen sink, and pass it along in hopes that something will eventually line up with data from other DFMs.**

Matching Logic Standard

In order to apply the matching logic consistently across different DFMs, all of the participants will have to calculate the variables and hashes in **exactly** the same way. The best way to do this is for DFMDC to publish a detailed standard through its Technical Working Group.

JSON Snippet	Purpose and Discussion
	The **Booking Report to DFMDC** contains the required matching hashes each for both load and truck: • A block of hashes for the hard IDs that are available for this load. Participating DFMs are tasked to query their customers for as many items as possible. Ideally, they would embed the requests in their API documentation. The DFMs probably won't get every ID type, but the more the better. • A block of hashes for ID elements that are "hardened" by combining several items before hashing. • A block of hashes for combinations of shipment metadata ("soft IDs") that help narrow possible duplication. • A block of text values that DFMDC can divert into its Data Lake to build general metrics about national trucking activity.
	DFMDC takes the Booking Reports it receives and forwards them as a stream of **Asset Usage Notifications** to the other DFMs. The notifications transmit much of the information that the Source DFM provided, minus the industry metadata and with the addition of calculations that DFMDC injects to help User DFMs weigh the notification's significance in their decision-making. DFMDC sees all of the "soft ID" hashes that were reported by all DFMs. From these, DFMDC can keep a running count of how often each appears in the truckload marketplace. These frequency counts are added and forwarded to the user DFMs.

<table>
<tr>
<td><pre>
{
    "hard_ID_hashes" : [
        {
        "GS1_ID" : "1645ce1ae74355830eaf2ee738d07177",
        "BITA_ID" : null,
        "EDI204_ID" : "3b412a8fe9da611e5fff48253074f36b",
        "Other_IDs" : null
        }
    ],
    "hardened_ID_hashes" : [
        {
        "oz_invoice_dz" : "4636f72e5a92c9a7c7ddb813ad20d2d1",
        "oz_pick_POnumber_dz" : "4636f72e5a92c9a7c7ddb813ad20d2d1"
        }
    ],
    "soft_ID_hashes" : [
        {
        "oz_pick_dz" : "57fd9244a785b98ac229718a912650d4",
        "oz_pick_tt_dz" : "f3422f3b1617339e6503106b0f2ff551",
        "other_soft" : null
        }
    ]
}
</pre></td>
<td>Each DFM is tasked to use the published ID Hashing Standard to construct a JSON data snippet that describes each asset (truck or load) as it is posted on their board.

Beyond simple formatting, each DFM will be tasked to make a best effort to draw as much matchable data from its contributors as possible.</td>
</tr>
<tr>
<td><pre>
{
"match_report" : [
    {
        "cust_dfm_ID" : "112",
        "dfmdc_msgID" : "9876543210",
        "load" : "true",
        "truck" : "false"
    }
    ]
}

{
“report_receipt" : [
    {
        "dfm_ID" : “45",
        “dfmdc_submit_no" : “909090909"
    }
    ]
}

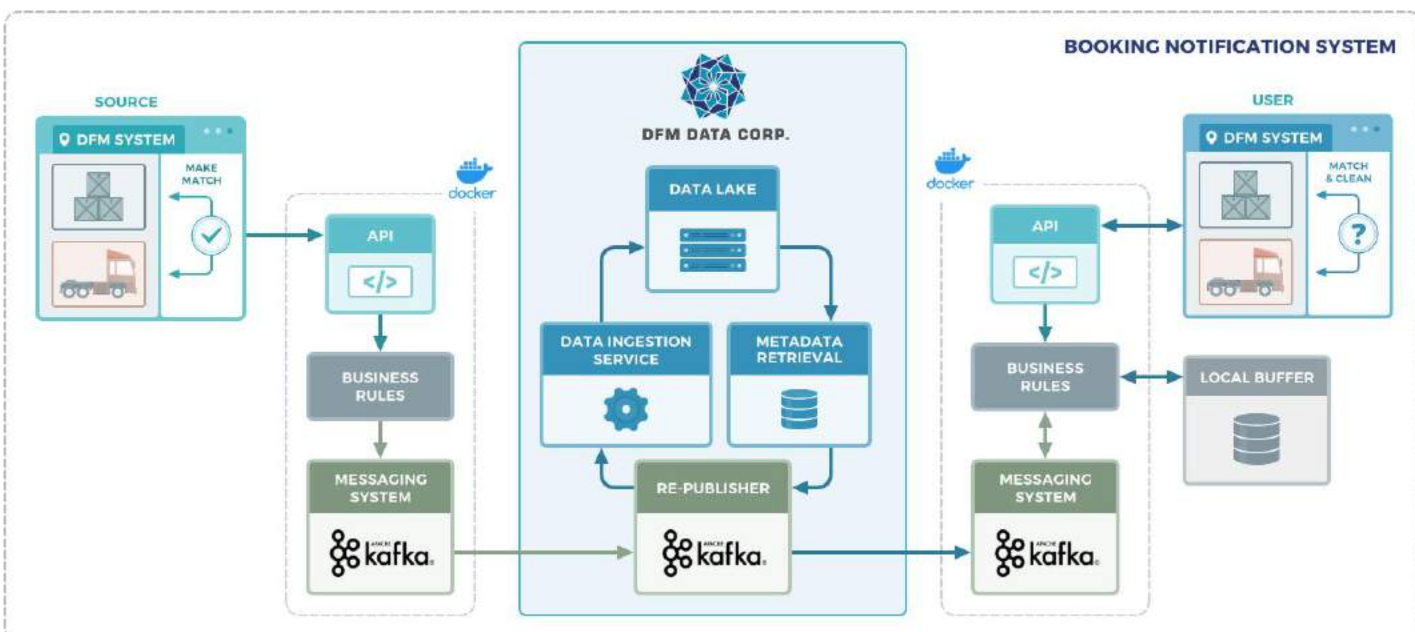
{
“duplication_match_report" : [
    {
        "dfm_ID" : "112",
        "load" : “4",
        "truck" : “0“,
        “dfmdc_submit_no " : “909090909“,
        “update_time” : “01292021-09:44”
    }
    ]
}
</pre></td>
<td>The system will exchange various short messages and acknowledgements to keep all of the participants informed about key performance parameters.

For example, if a User DFM identifies a match, a report would be sent back to DFMDC indicating that the match had occurred. No additional details would be sent, but this message would allow DFMDC to build a picture of how often duplication is happening and possibly narrow down the places where it is most common.</td>
</tr>
</table>

Table 3 – Notification Data

Notification System Infrastructure

To use the matching system in practice, DFMDC has built a cloud-based system to collect, distribute and compare the large number of IDs that the US trucking industry generates each day. The system is deployed in two key components. One is a package of software that is assembled, distributed and maintained in a Docker container. This software suite contains the functions to input booking reports and send them to DFMDC's central Data Lake and re-publisher. The suite also contains the code required to receive a large stream of JSON/YAML messages and compare them to reference JSON snippets constructed from each members DFM's active listings.



24-Feb-21

Figure 6 - Notification System Infrastructure

The reason for using Docker containers (or their equivalent) is to construct a self-contained package of code that each participating member DFM can install and use as they see fit. Depending on the needs, they could be installed and hosted in DFMDC's cloud (for a fee), hosted in a member DFM's corporate cloud or even installed and run on DFM's premises. Updating the software should be as easy as cloning the latest Docker version from DFMDC.

The messaging would be coordinated through a central software system hosted in a cloud that is managed by DFMDC.

Members of the DFMDC Consortium would be invited to view, comment and contribute to the software in both the distributed Docker installations and the central system. Likely implementations would rely on readable, interpreted scripting languages (e.g., JavaScript, Kotlin and/or Python) and viewing and development would occur in GitHub or equivalent. The software would be owned by the consortium, and the development and maintenance process would resemble an open-source project. All Consortium members would be invited collaborators.

The core matching process will happen in the software contained in the standard Docker Container. Figure 6 illustrates one plausible way to test for matches. In this method, the DFMDC Docker container is installed by each member DFM, along with a database or repository that resides outside the container (so it is persistent if the Docker container is updated). The database contains a buffer of all of the notification messages that the member DFM has received from DFMDC in the recent past. The database also contains a buffer with a list of all the truck and load postings that the member DFM has accepted in its marketplace. Both buffers (from DFMDC and locally generated) are first-in-first-out and sized by the member DFM (e.g., 24hrs, 72hrs, 96hrs, etc.).

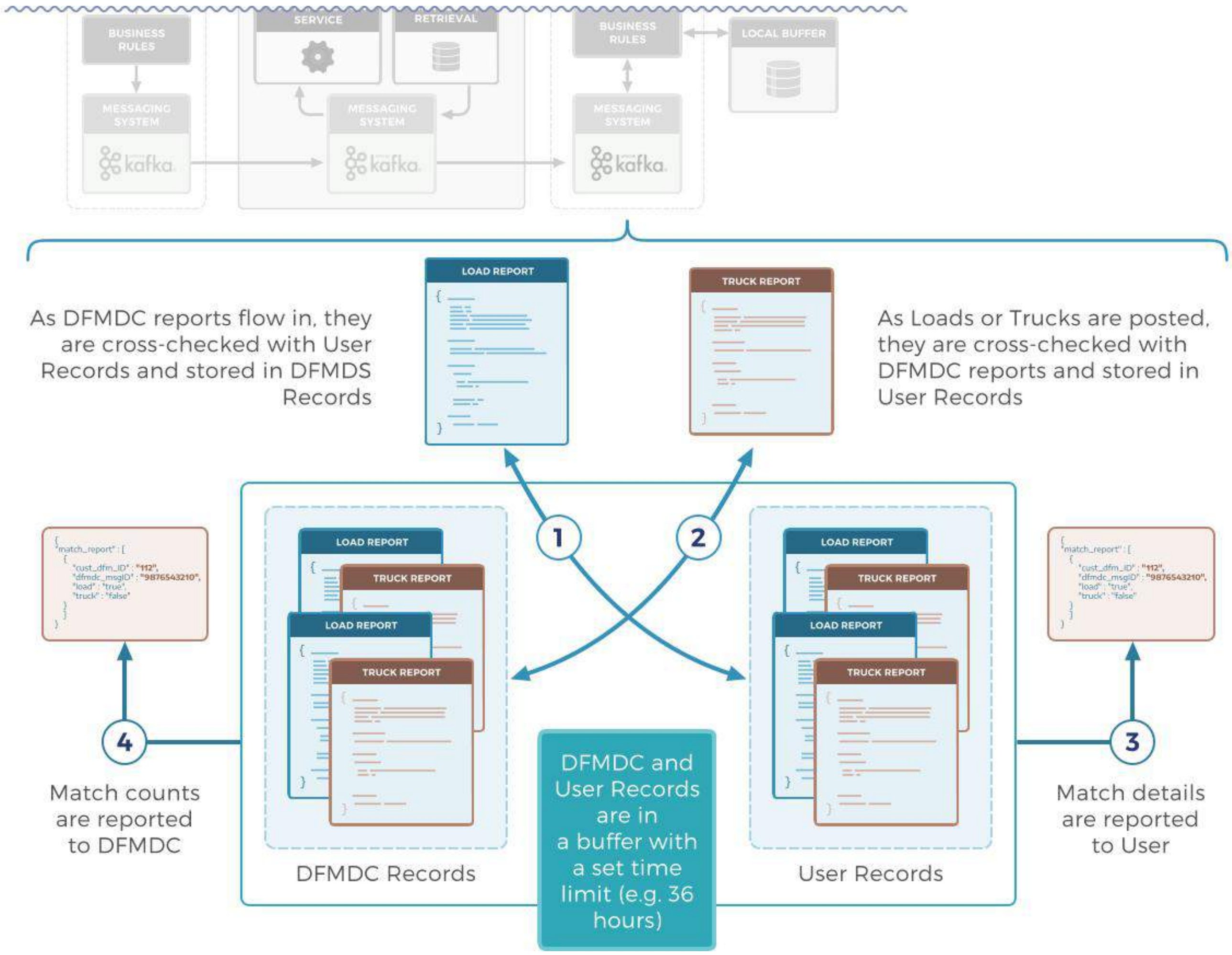


Figure 7 – A Possible Matching System Design

With the two buffered lists of load and truck reports, the workflow could be as follows:

1. As each notification is received from DFMDC, the load and truck information are compared to the buffered reports from the member DFM.
2. As each new listing occurs, the user report is compared to the notifications received from DFMDC.
3. If there is a match, the details are reported to the member system, and
4. The match/no-match results for load and truck are reported to the DFMDC.

In this system, a comparison occurs twice (one for the load and one for the truck) when a new notification report comes in from the DFMDC and also when each new listing report is inserted by the member DFM.

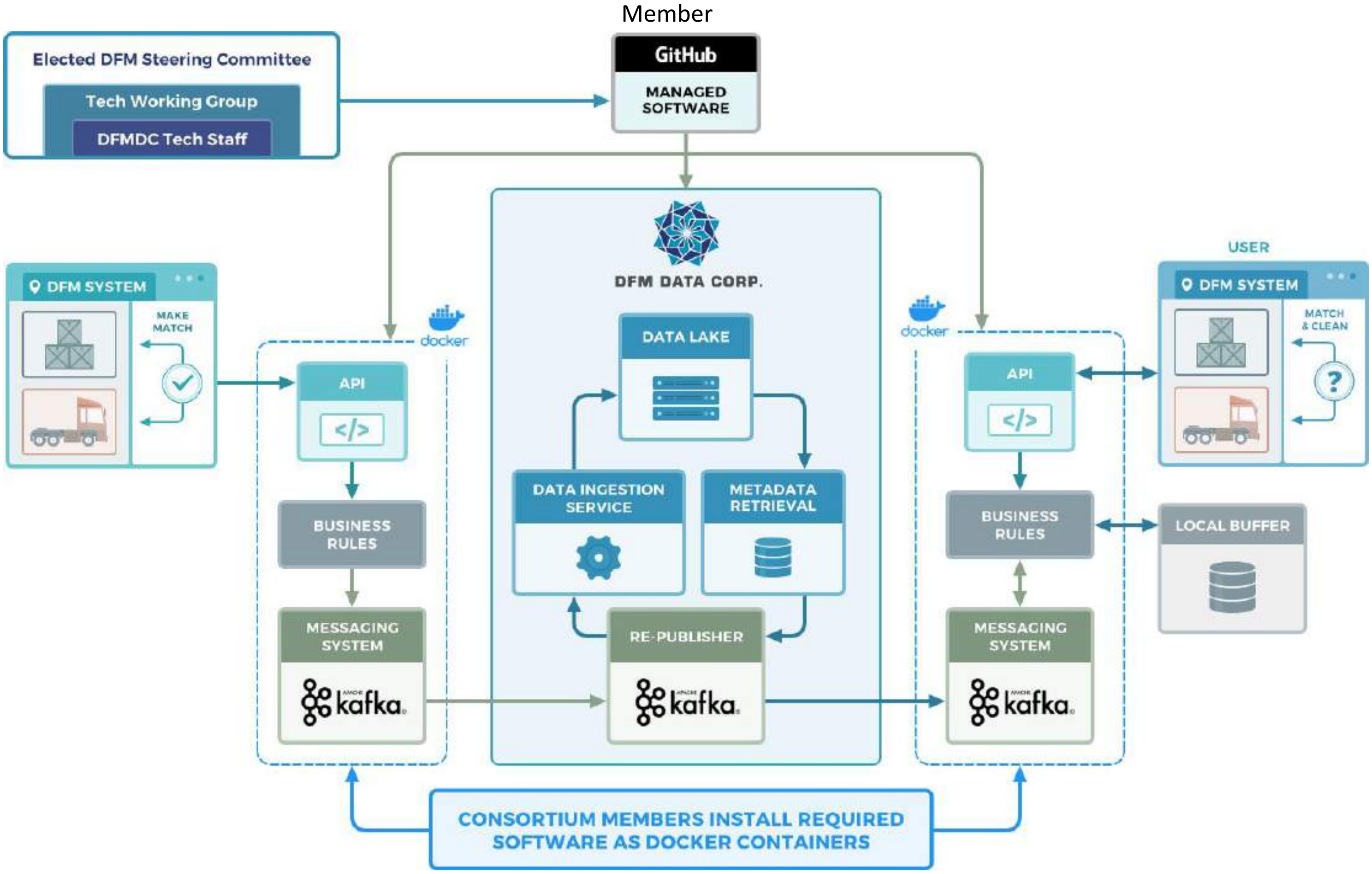


Figure 8 - Software Development and Maintenance

Giving Consortium members access to the source code and packaging everything in a Docker container will allow the DFMDC Technical Working Group to maintain control and consistency over a widely distributed system. It will also reassure member DFMs that their data is being handled safely. For example, it may make sense for the member DFMs to submit their data in clear text and let the Docker software calculate the hashes. That could create risks, but having the code open and visible would reassure member DFMs that their input data are not being retained or passed on once the hashes are calculated. Similar assurances can be provided in every other part of the system.

System Message Volumes

The volume of messages in the DFMDC system will depend on the number and size of the member DFMs that participate. The basic message structure is shown in Figure 9. Each source DFM will contribute a booking report whenever a load and truck are matched. DFMDC will take each booking report and split it into two streams (truck and load separately) and re-transmit the stream to every other member DFM. If there are 100 DFMs in the Consortium, then the collective daily report input from all 100 members will be rebroadcast 198 times. Each member DFM will receive a copy of the daily report volume, and the software in the Docker container must evaluate comparisons from all those notices each day.

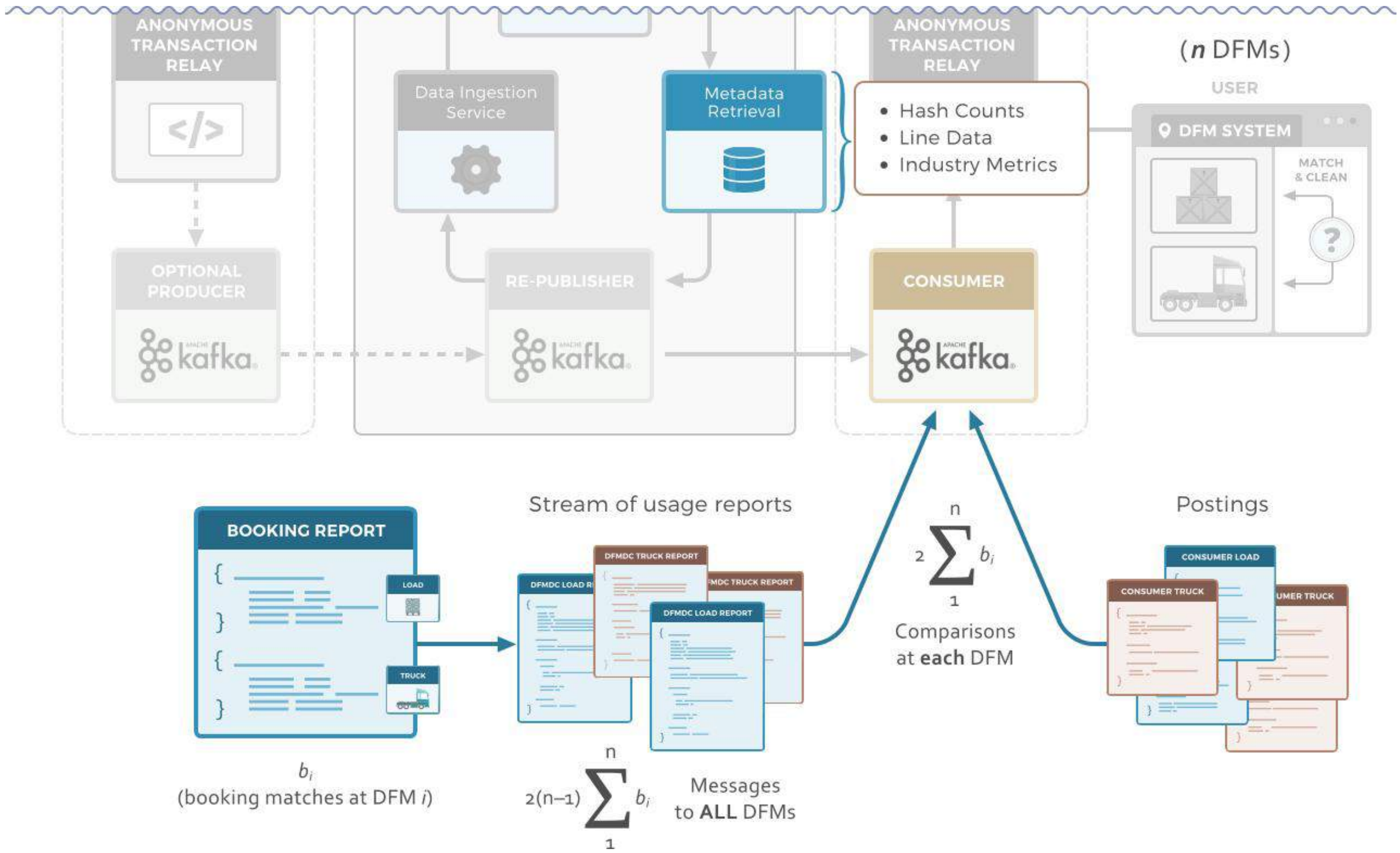


Figure 9 - Message Volume

It is not hard to imagine that the system could be asked to distribute up to 200 million messages per day.

In addition, there will be a number of smaller acknowledgement messages that might increase that number several times.

Finally, these numbers assume that Source DFMs only issue one booking report for each truck/load combination. However, bookings actually occur in stages, with varying levels of legal validity. It may make sense to issue booking reports when the bookings are made, when the truck arrives, when the load has left, or when the Bill of Lading is finalized/signed by recipient. Possibly each might be reported as a distinct “event” using the same matching logic.

Taken together, it is at least possible that the system may be asked to handle a billion messages per day.

Notification System Metrics

A lot of messages will pass through the system. The following diagram shows the data flows through additional DFMDC processing and ties back to the numbers of member DFMs and loads booked in the US trucking market.

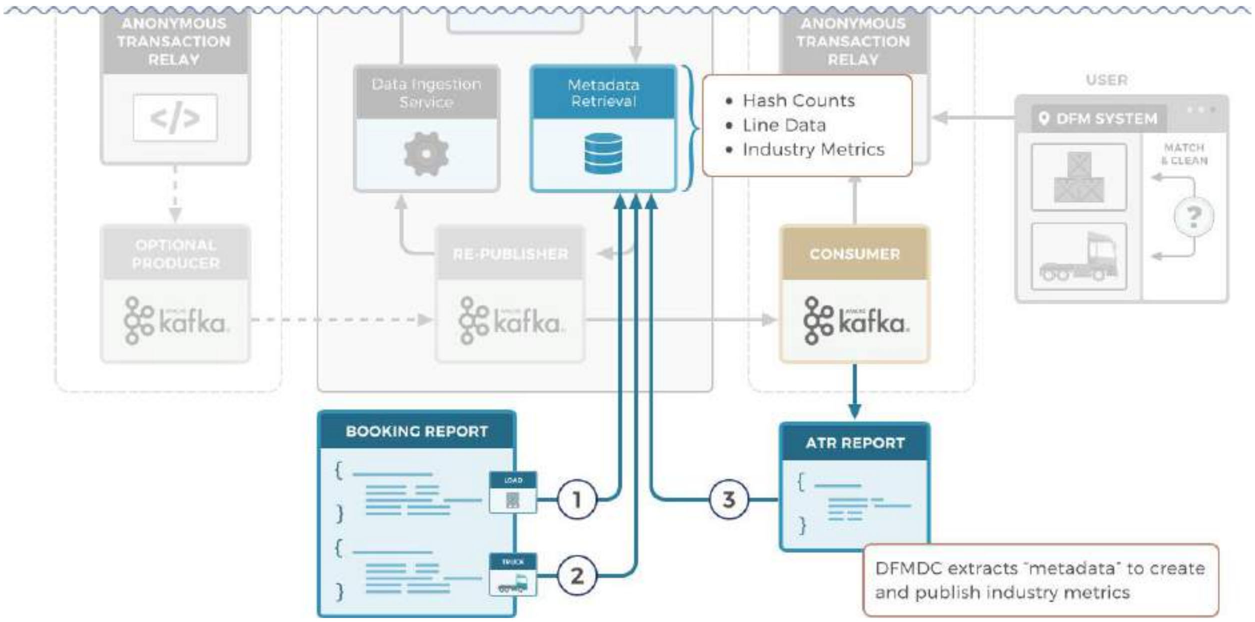


Figure 10 - An Engine for Industry Metrics?

1. The DFMDC strips out the clear text variables and stores those in the Data Lake. It sends copies of all the hashed values to all of the member DFMs, along with a count of how often each hashed value has appeared in recent booking notifications.
2. If a member DFM computes a soft ID hash that matches a hash received from the DFMDC, it can see the frequency count and know how often that hash has appeared in, say, the past 24 hours. For example, if there was only one listing for a given load, there is a very good chance that the load listed on the members DFM's load-board is the same load. This is not decisive, but it may cause the member DFM to place a warning badge on that listing on its load-board.
3. Member DFMs can subscribe to view aggregated summaries of the collected data and the associated analysis results.

In theory, it would be possible to collect a wide variety of text metadata from the booking reports. These data items would not be passed to Member DFMs for matching and data cleaning purposes, but they would go into the DFMDC Data Lake and become available for computation into aggregated industry metrics. For example, origin and destination zip codes might be included for each booked load. DFMDC's Data Lake would ingest a continuous flow of near-real-time lane data for the US spot truckload market. How or if that data should be shared would be a subject for the Technical Working Group and the DFMDC Steering Committee.

Quality Control

One area for possible concern is the worry that some participating DFM members might, out of error or self-interest, attempt to seed the system with erroneous data. The data in the DFMDC Data Lake will provide at least a few tools to protect the system from flawed behavior.

For example, suppose that member DFM A does not submit Booking Reports for all matches. It does, however, consume and test every Usage Notification. On the surface, DFM A might seem to be gaining benefit without contributing to the community.

However, this situation would create an imbalance that should be evident in the Data Lake. If member DFM A reports match against Usage Notifications from member DFM B, but member DFM B reports no matches against Usage Notifications from member DFM A, that should be flagged as an unlikely and anomalous situation. At the least, it would justify an audit of DFM A's installation.

We anticipate that various quality checks and remedies will emerge as everyone gains experience with the system. Some may even cross over to become industry metrics that are worthy of reporting to Consortium members.

Standards Development

While we can sketch out the general principles of the matching system, it will take a lot of experimentation and refinement to make it efficient and effective. Ultimately, the load matching criteria must be a consensus design by the DFM community. Participating DFMs will need to make changes to their business systems to collect and organize the necessary input data to pass to DFMDC to construct the required hashes. Getting there from here will require a process that is outlined below.

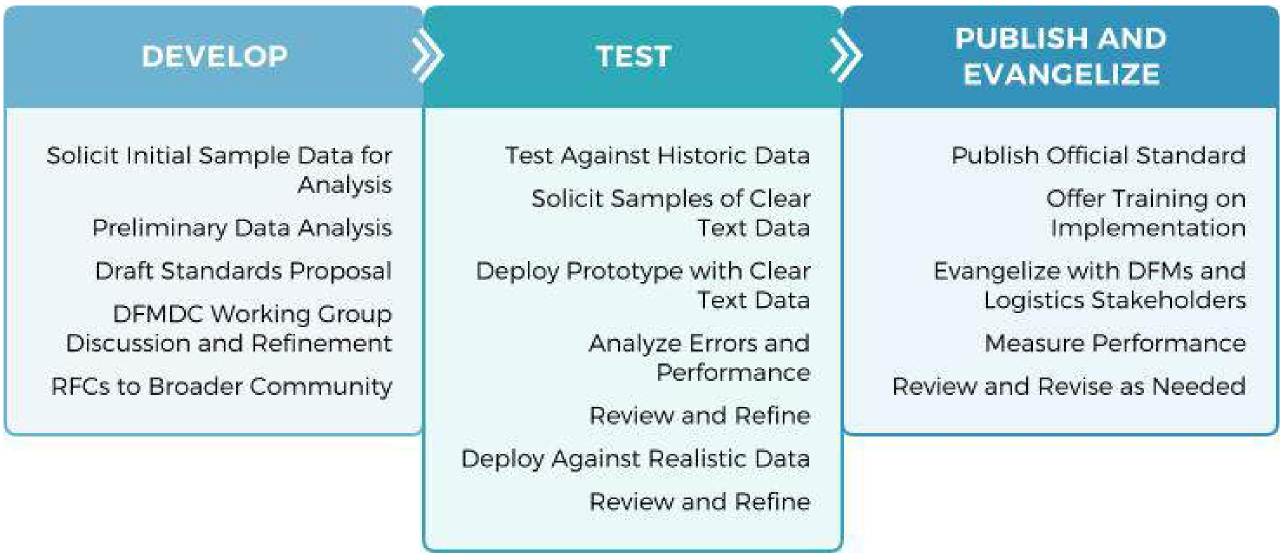


Figure 11 – Tentative Standards Development Process

At a high level, the standards development process involves several steps:

- Confer with an initial group of DFMs and contributing members to identify potential data that are candidates for use as matching criteria.
- Solicit one or more samples of typical load posting data from DFMs. Use the data to analyze the performance of the proposed criteria and identify what changes are required to create matchable reports.
- Publish a draft standard as a “request for comments” (RFC) outlining the initial suggested variables and share the RFC with interested stakeholders. This will ask participants about the availability and practicality of using the proposed matching IDs.
- From the responses, refine and publish a draft data collection standard and evangelize it with the participating DFMs as well as with DFMs that may be considering joining the Consortium.

The immediate goal is to build a system of matchable reports that can be adopted by the industry as a whole. That will be an iterative and incremental process. The long-term goal is to build industry consensus to adopt standard data schemas for efficient data cleaning industry wide. Ultimately, that may connect to other standardization initiatives like EDI or GS1 that might supply unique matchable IDs to drive the DFMDC process.

Appendix 1 – Considerations for Matchable Shipment Data

Different Types of Matching Error

The goal of DFMDC's phantom data cleaning system is to recognize when a load or truck on one DFM is the same physical load or truck that is listed on another DFM. In this matching process, there are three possible outcomes:

- An accurate match
- A false positive – An apparent match is found, but it is a mistake. No real match exists.
- A false negative – A real match exists, but the matching process misses it.

False positives are more worrisome than false negatives. A false positive could cause a DFM to remove a valid listing from its board. That would hurt the DFM's client and reflect badly on the DFM's reputation.

False negatives are undesirable, but not as threatening. They mean that some phantom data will be missed in the cleaning process. In the initial development stage of the system, that will happen anyway. It would be ideal to clean 100% of the incorrect listings, but eliminating 80% or 90% would still be a big win for the industry and a basis for future improvement.

Data must be consistently entered

In order to get a match between listings on different DFMs, the hashes must match. That means the data that is input to the hashes must be absolutely identical.

DFMDC must publish very precise standards on how each ID should be constructed and the DFMs must work to solicit information from their customers in as consistent a manner as possible.

Some types of data will be more suitable than others for constructing matchable hashes. For example, phone numbers are often written in different formats:

- (333)-123-4568
- 333.123.4568
- 333-123 4568

However, it is easy to make phone numbers matchable by stripping all the extraneous characters and punctuation and formatting them as a simple number: 3331234568. Software can be applied to coerce everything to the simple form, likely when DFM records are inserted into DFMDC's program for comparison.

Other possible identifications are harder to make consistent. Names, for example can be written in so many different ways that they are poor candidates for matching. Consider:

- Jason Williams
- Jase Williams
- J.A. Williams
- Jay Williams

These might be the same person. Some might be nicknames. Some might be different people. It is extremely hard to tell. In addition, if these were hashed, the results would be very different and there would be no way to determine if they were even similar, let alone identical.

The bottom line is that DFMDC and its DFM partners have to search for matchable identifiers that DFM clients are likely to enter in a consistent manner. Then, if possible, those entries need to be coerced into rigid formats that allow even less room for variation.

The variables must come from one source

In order for truck, load and booking records to be comparable, the ID values must come from the original physical asset that is being tracked. A DFM can assign an identifier to its listings, but we need ID information that is exactly the same in every DFM where the load or truck is listed. That can only come from the source of the actual load or truck.

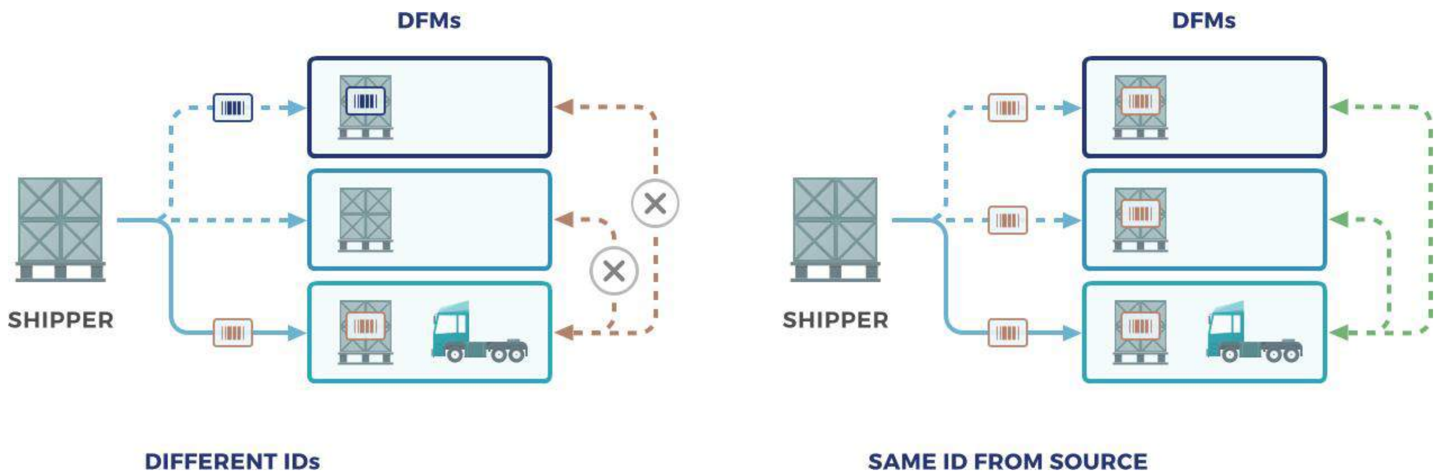


Figure 12 - IDs must originate at the source

Any strong ID is enough as long as it is used consistently

If we can find a strongly unique ID, the resulting match should prove that the listed loads or trucks are identical. We don't need to use the same type of strong ID for every load or truck, as long as the same one is used in every DFM listing for that truck or load.

For example, if all DFMs solicited a "company ID" in their listings, presumably a given shipper will submit the same code to every DFM. The origin of the code could differ. One shipper might supply SAP sales number. Another supply an invoice number. The key is that the supplied ID is unique across all shipments by that customer. It doesn't have to be the same type of code used by a different shipper.

The supplied code can be combined with other variables to make the input more distinctive. For example:

- Shipper invoice number: "I27309"
- Load origin zip code: "30303"
- Load destination zip code: "95014"

The input to hashing might be: "30303+I27309+95014"

If a different shipper used a TMS to post to multiple boards, its TMS might be coded to insert the same identifier into every posting for a given load. The hash structure might be different:

- TMS-supplied ID: 80990990

All DFMs would hash the ID from the same input: "30312+80990990+95080"

The inputs to the two hash IDs come from totally different sources, but it each one is the same source for all postings for that shipment from that shipper. If that is preserved in every participating DFM, the resulting hashes will be matchable across all participating DFMs.

Imprecise data can be made more consistent by using "bins"

To be useful in hashing, a data variable must be entered the same way in every DFM that hopes to remove duplicates. Some variables like phone numbers are consistent if the punctuation is stripped away. However, many numeric values are likely to be entered with small differences that are not meaningful to shippers and truckers but will create different hash outputs. For example, if asked to estimate deadhead miles for a load, a shipper may enter 70 miles in one DFM, 70.0 miles in another and 75 miles in a third. If these slightly different numbers were incorporated into a hash, matching would be impossible.

One way to reduce this problem is to define deadhead miles in terms of a standard set of "bins" that can be offered as drop-down or API choices when shippers are entering the data for their loads. For example:

- 0 to 20
- 21 to 40
- 41 to 60
- 61 to 100
- etc.

There is an implicit tradeoff with this approach. The bins are not as precise and will not discriminate between listings as effectively as entered numbers. On the other hand, bins are more likely to capture the intended values and less likely to generate false negatives.

Correlated variables are useless

The following table applies three different soft ID hash constructions to a sample of 30,273 loads that were supplied by cooperating DFMs. The way to read the table is to look at the percentage and the number of loads that occurred with that specific hash input. Hence the value 59.6% means that 59.6% of all loads involved an "originZip+destinationZip" combination that occurred just once. That combination occurred twice on 11.3% of the loads. When the shipment weight (rounded or binned to the next largest 1000 lbs.) is included in the zip, the number of singletons increases to 84%.

Number of Loads that Recur	Percentage of Loads with # of Occurrences by Inputs to Hash Construction		
	oZip+dZip	oZip+wt+dZip	oZip+miles+dZip
1	59.6%	84.2%	59.9%
2	11.3%	6.7%	11.3%
3	5.5%	2.5%	5.5%
4	4.2%	1.3%	4.2%
5	2.7%	0.7%	2.6%
6	2.2%	0.7%	2.2%
7	1.6%	0.4%	1.4%
8	1.5%	0.5%	1.5%

Table 4 - Frequency of Occurrence of Different Hash Construction Inputs

When the shipment weight (rounded or binned to the nearest 1000 lbs.) is included in the zip, the number of singletons increases to 84%. However, if the distance in miles is used instead of weight, there is almost no improvement over origin and destination zips in terms of discriminating between different loads. The reason for the difference is that distance is highly correlated with the combination of zips, while shipment weight is not. Adding distance to the hash is pointless.

CAVEAT: On its face, this appears to suggest that we could distinguish (and therefore match) 84% of all listings if DFMs would provide DFMDC with the origin and destination zips and the shipping weight. While this is true for this sample of 30,000 listings, it would not be true if DFMDC were handling 500,000 or a million listings per day. In that case, the number of repetitions of hashes would increase dramatically. In order to understand how badly the matching performance would degrade, we need to collect and test a much bigger sample of typical data.

DFM DATA CORP.

Interconnected. Interoperable. Invaluable.

Request for Comment

Draft Method for a Transport Unit IDentifier (TUID)

Problem Statement

As shown in Figure 1, when a shipment happens, it generally progresses through a shipment life cycle **(1)** with two major phases. The physical Transport phase is generally familiar and well-understood. There are many formal characteristics that are currently standardized and automatable once a load is prepared for shipping and/or a Carrier is booked. They allow industry procedures and software to know and track each load as it makes its journey.

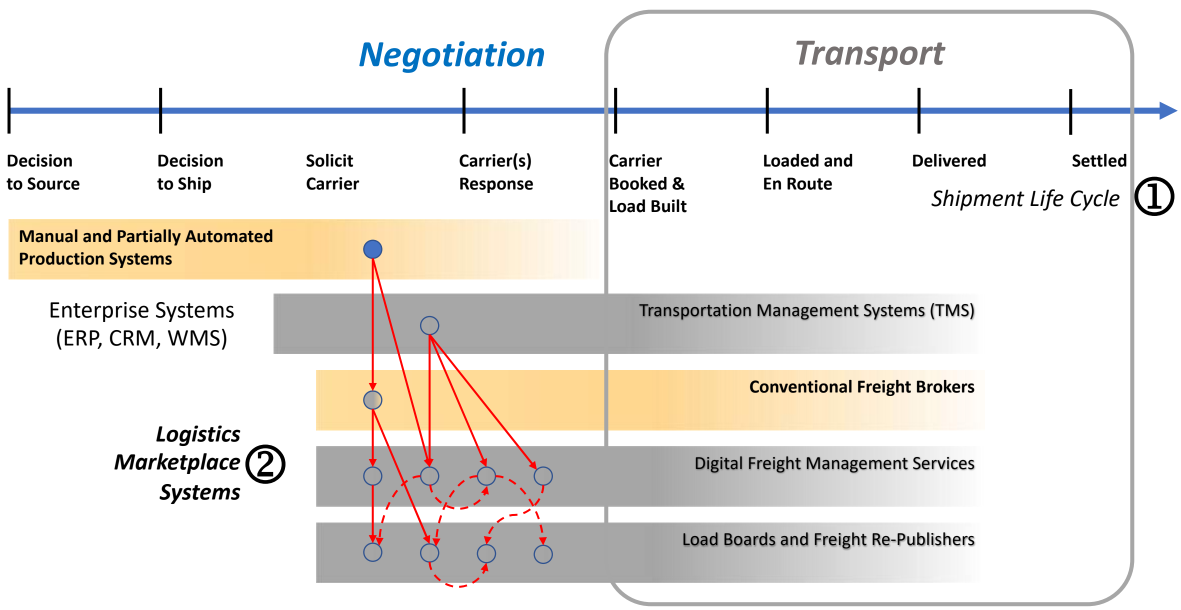


Figure 1 – Negotiation and Execution Phases in Trucking

However, before the load is built and the Carrier is booked, industry participants often engage in **complex negotiations** to match loads and carriers. These may encompass familiar and new parties across the industry, especially if the shipper elects to explore the Spot Market **(2).** Until recently, the negotiation phase in the North American trucking industry relied predominantly on human-centric procedures and systems. Now there is growing urgency to bring modern digital procedures and systems to this activity. One big roadblock is the fact that there is no widely accepted way to identify and label loads while they are still in the negotiation stage. It is hard to automate things that don't have clear labels or names.

Appendix 1 gives more detail on key differences between the Negotiation and Transport phases of the shipment life cycle.

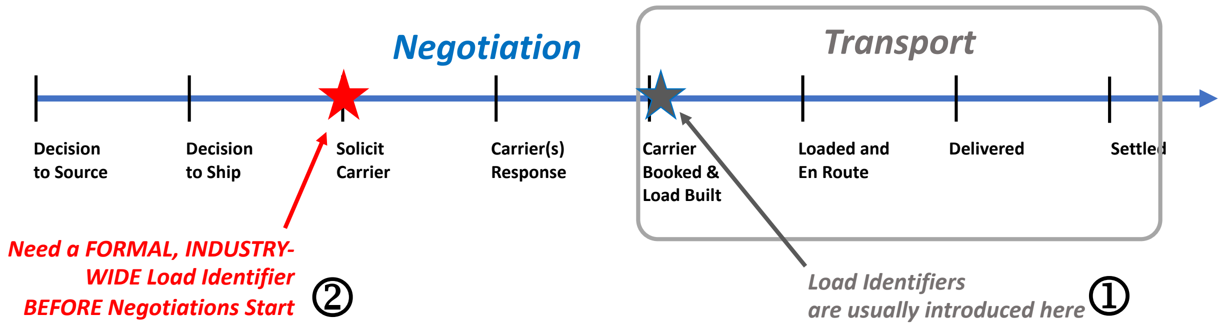


Figure 2 - Where Unique Identifiers Are Created

Need for Unique Load Identifiers in Negotiation Phase

At present, automatable load identifiers usually emerge at the start of the Transport phase, when a load is prepared for shipping and/or a Carrier is booked **(1)**. These identifiers may be EDI numbers, RFID codes, barcodes, shipping manifest numbers or any number of other forms. They exist so industry systems and software can know which load is which and possibly where it is located at a given moment. It is common practice to use these identifiers in conjunction with formal labels (barcode, BOL number, RFID code, EDI number, etc.) that support automation.

Unfortunately, there is no comparable load identification method that applies widely during the negotiation phase. This wasn't a huge problem when negotiations were human-to-human, but it's a serious roadblock to large-scale digitization of the market-making process. Even the best AI will struggle when a load is described as "the 38,000 lb. of wire coil at Gate 3 on our loading dock." DFMDC believes the industry can clear this roadblock if it implements the simple innovation depicted in Figure 2.

DFMDC proposes that Shippers unilaterally generate unique identifiers for their loads as soon as they form the clear intent to make the shipment **(2)**. We propose a method (detailed in Appendix 2) to generate a uniform **"transport unit identifier" (TUID)** that can be constructed by each shipper independently and shared with the intended shipment Recipient, and any intermediaries (Carriers, Brokers and DFMs) that participate in the negotiations.

If that TUID existed, systems and software could leverage it to reference, track, describe and record both **prospective and actual** shipments. This would make it easier to create digitized, value-added activities and services for the negotiation process. Achieving such new value-added activities will require buy-in by all of the affected parties … but a common TUID would at least make them possible.

TUID Generation

To be effective, the TUID doesn't need to be intrinsically meaningful. It just needs to be unique across all shipments from all Shippers. There are many methods to achieve this, but this RFC outlines a method that is simple and requires minimal technology and industry coordination.

The details are expanded on in Appendix 2, but the essence of the idea is as follows:

- Leverage simple identifiers that are readily available in every organization such as phone country codes, company tax ID numbers and whatever internal shipping or accounting numbers they normally assign to distinguish their planned loads.
- Assemble text representations of these identifiers according to a strict construction method.
- Hash the resulting text string to create a load identifier that is unique, but otherwise meaningless. It is a label, nothing more.

A TUID of this type would allow each shipper to label their intended shipments from the beginning of the shipment life cycle in a clear, unique and secure manner.

Application and Deployment Strategy

If a TUID method is defined, the larger challenge is to entice the myriad participants in the trucking industry to adopt it as a standard data element in their negotiations and deliberations. FWIW, getting the trucking industry to agree to follow a single practice probably makes herding cats seem like a kindergarten exercise. DFMDC's suggestion for an implementation strategy involves two related initiatives:

- Establish an organization that can institute and administer a neutral governance system where industry stakeholders can join, engage and ultimately control the standardization process.
- Create a roadmap (implemented through the governance system) to introduce the TUID in a way that will entice participation rather than attempt to mandate or enforce it. The most plausible way is to create incentives for the ultimate customers (and funding sources) to demand the TUID from the industry.

There is no obvious priority or order to these activities since both are essential to success. The following sections give more ideas on how they might work.

Establish a Neutral Governance System

Most successful standards are supported by a formal governance system to manage definition, publication, enhancement, and (sometimes) enforcement. For more detail, see Appendix 3. In the simple case of load identification, the goal of a governance system is to give industry participants confidence that the standard is neutral and will be applied in a way that is true to its stated purpose.

FWIW, the other way to achieve standardization is for a government to mandate a standard and enforce standards compliance. DFMDC's ideas on governance are motivated by the belief that the US trucking industry would prefer to avoid the government mandate approach.

As a legal entity, DFMDC is explicitly designed to act as a governance system to apply the neutral standards and systems administration tools this problem requires:

- DFMDC is a "for benefit" legal entity with bylaws that direct it to donate 25% of all profits to a charitable cause.
- Currently, DFMDC has 140 equity shares set aside for sale to active industry members. These shares give the owner the right to one vote in the selection of DFMDC's 8-seat steering committee.

- The majority of the seats on the Steering Committee are reserved for industry participants. When sufficiently subscribed, the industry participants will control the Steering Committee that drives the policies of DFMDC.
- The Steering committee would have explicit control over the roadmap for DFMDC service offerings, technology choices and the subscriptions and user fees that will be required to fund DFMDC's operations.

The intent of this governance system design is to place the industry stakeholders, through a democratic process, in direct control of DFMDC's funding, actions and operations. In other words, the industry participants will run the show.

Enact a Voluntary Industry Implementation Strategy

A TUID can be generated easily and independently, either unilaterally by the shipper or supplied from a centralized service. The key is the fact that it relies heavily on pre-existing industry identifiers that are already designed to be unique. Any company or shipping entity can construct its own TUIDs and be confident that they will not collide with others.

The more daunting challenge is to devise a strategy to entice the industry to adopt the TUID on a broad enough basis to be effective. Asking a government (arguably the US Federal Government) to mandate its use is one way to achieve this, but probably not the preferred approach for many members of the industry. DFMDC has identified two complementary scenarios that might drive widespread adoption by voluntary industry means.

Option 1 - Leverage Enterprise Software

The first approach is shown in Figure 3. In this approach, the makers of enterprise software systems such as Enterprise Resource Planning (ERP) and Transportation Management Systems (TMS) are encouraged to add an option to generate a compliant TUID at the appropriate point in the shipping process.

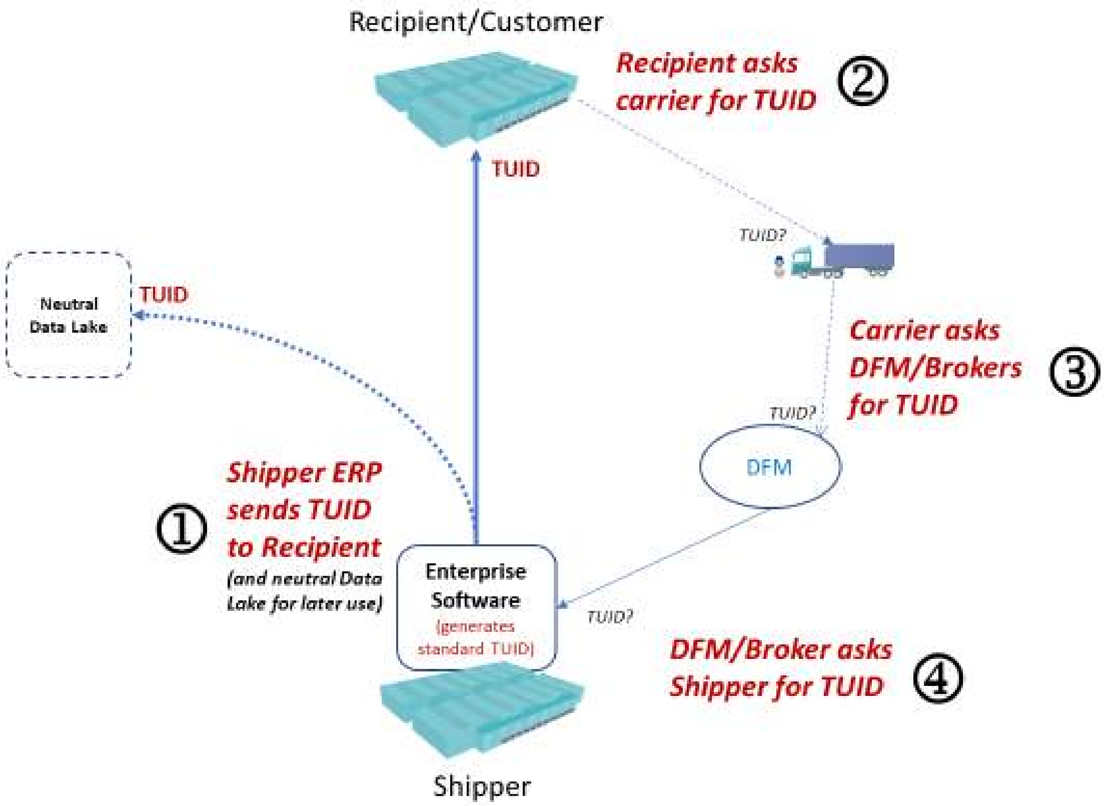


Figure 3 - TUID from Enterprise Software

1. At a point **prior** to the time when the Shipper begins to solicit a Carrier for the load, the enterprise system generates a standard TUID and sends it to the intended shipment Recipient.
2. The Recipient is encouraged to note the Shipper-supplied TUID and demand that a matching IDs be supplied by the Carrier when the goods are delivered.
3. That will encourage the Carrier to demand the TUID when a Broker or DFM books it to carry the load.
4. That, in turn, will encourage the DFMs and Brokers to demand that Shippers supply a valid TUID and that will encourage Shippers to apply their enterprise software to generate it automatically.

The TUID generation and sharing process is intentionally simple and stand-alone. There is no intent to control or restrict who can perform the generation process. An Excel spreadsheet can be programmed to do it. However, it is assumed that the process will be more consistent and credible if it is embedded in widely accepted enterprise software (e.g., SAP, Oracle, Manhattan, Salesforce, or McLeod).

Option 2 - TUID Generated by Public Server

Some Shippers may not have a suitable enterprise system. A software service can be established (Figure 4) to input the required information and generate the required TUID.

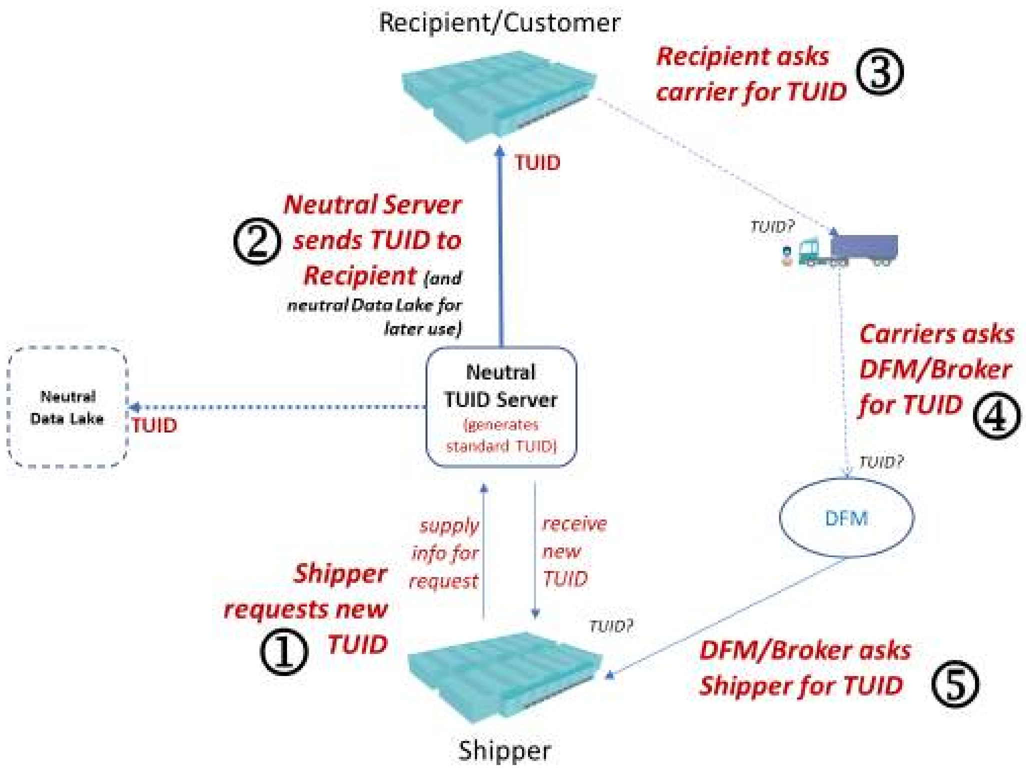


Figure 4 - Stand Alone TUID Generation

1. At a point **prior** to the time when the Shipper goes out the market to solicit a Carrier for the load, the Shipper submits the necessary data to generate a compliant TUID
2. The software service returns the TUID to the Shipper and sends it to the intended Recipient.
3. The Recipient is encouraged to note the Shipper-supplied TUID and demand that a matching TUID be supplied by the Carrier when the goods are delivered.
4. That will encourage the Carrier to demand the TUID when a Broker or DFM books it to carry the load.
5. That, in turn, will encourage the DFMs and Brokers to demand that Shippers supply a valid TUID and that will encourage Shippers to apply their enterprise software to generate it automatically.

(Optional) Neutral Data Lake

It is essential that the TUID be sent to the intended Recipient of the shipment. This is a cross-check to reduce the risk that the TUID will be generated and used frivolously or fraudulently. A shipper might consider issuing multiple TUIDs to hide its actions in the marketplace, but it is less likely to do so if each issuance tells its customer that another load is enroute.

This simple safeguard can be greatly strengthened if the TUID is also sent **to a Neutral Data Lake.** If each TUID instance were recorded in a common Industry Data Lake, there would be an authoritative reference point to verify TUID authenticity and, possibly, collect industry data that might offer useful opportunities to improve industry operations. For example:

- Brokers might be able to look up a submitted TUID to verify authenticity.
- Additional descriptive data could be added to the Data Lake (linked to the TUID) to build an industry neutral picture of trucking operations. This could generate better lane data and cleaner industry records.
- Obsolete, abandoned and already-booked loads could be cleaned more efficiently from industry listings and operations.
- etc.

However, once we begin to consider the collection of additional data and speculate on possible ways to use it, a host of ethical, competitive, and operational questions arise. What data will be collected? Who will be able to see it? What additional services will be constructed?

If systems like these are to be developed on a voluntary industry basis, the effort must be controlled and guided by a governance system that (at least most) industry stakeholders can trust. This brings us back to the need for the DFMDC governance system outlined above and in Appendix 3. Systems like these will only succeed if they can be trusted to be fair and transparent.

Overcoming Resistance

In the US Trucking Industry, any proposal of this type will encounter strong resistance. To overcome that resistance, DFMDC's TUID initiative is looking for partners and collaborators that can help to remove some of the fear and corroborate the benefits that a better-defined negotiation process could offer.

DFMDC is exploring several ways to incentivize adoption and reduce resistance:

- Since the TUID mechanism is so simple, the cost of TUID generation can be minimal or even free to encourage widespread adoption of the system. The proposed governance system provides a vehicle to set up a fair and transparent funding method.
- DFMDC and the governance participants can articulate other ways that TUIDs can be leveraged by industry participants to secure and streamline their operations. These are more likely to involve paid services. There is a strong possibility that innovators and entrepreneurs will be able to make good use of the TUIDs to underpin totally new types of services. All of this will need a community effort, both to ensure credibility and to speed acceptance.
- The proposed deployment is patterned as a "customer pull" approach. If shipment recipients (aka customers) can be convinced that they should demand TUIDs and they will demand that the intermediaries supply them on inbound shipments. The intermediary community may resist efforts to involve them in the TUID process, but that resistance will be greatly reduced if it is a customer demand.

Altogether, DFMDC believes that a few strategic influencers may be enough to jump-start this initiative. If a few big Shippers, ERP makers and customers were to signal their buy-in, the rest of the industry would likely find it easiest to adopt it for all shipments.

Summary

DFMDC proposes that the US (and possibly world) trucking industry adopt a standard method to designate intended shipment loads at their first inception. In its simplest form, the proposed TUID would be a unique, but otherwise meaningless, label that is created and applied to each intended shipment. If the other industry participants agreed to accept, use and forward this TUID in their negotiations, it would be possible to be far more precise about the state of each intended and actual shipment.

To implement this method, DFMDC has outlined three essential components:

- A proposed construction method for an industry-unique Transport Unit Identifier (TUID). This can used by anyone to create a TUID that will not conflict with any other that is properly constructed.
- A neutral, transparent, governance system to engage industry stakeholders in the process of refining and expanding this concept to apply industry wide.
- A roadmap to incentivize and encourage industry participants to adopt the method in their operations.

This request for comment is intended to put this idea in front of industry stakeholders as a discussion starter and to solicit input, ideas, and hopefully, support to try to make it (or something similar) happen.

NOTE: If you want to join the conversation, contribute your suggestions, and let us know that you would like to be updated on industry feedback, please email your thoughts to comment@DFMData.com. Thanks for helping us make this small step … it could lead to a giant leap forward in digitization!

Appendix 1

Negotiation Phase vs Transport Phase Comparison

The shipment life cycle can be split into two phases:

- **The Negotiation Phase** – Many participants engage in market-making to match proposed shipments with the carriers required to deliver them.
- **The Transport Phase** – The contracting parties do the work required to complete the delivery.

It seems that many of the recent innovations in trucking (e.g., load tracking, fuel efficiency, driver work rules) have been concentrated in the Transport Phase. That is not unreasonable because it is only at the start of the Transport Phase that the actual participants are known. By then, the challenge only involves the shipper, carrier, recipient, and maybe the broker or 3PL.

By contrast, the negotiation phase can involve many participants. Not all of them will be involved in a given shipment, but many more might become involved if circumstances change.

Negotiation Participants	Transport Participants
• **Shippers (hundreds of thousands)** • **Brokers (15,000 plus in US)** • **Digital Freight Matching Cos. (250+ in US)** • **Carriers (hundreds of thousands)** • **Recipients/Customers (millions)**	• 1 Shipper • 1 Carrier • 1 Recipient/Customer

The sheer number of possible participants in the Negotiation Phase makes it harder to implement automated systems. There are too many actors, too many agendas and too many competing initiatives. Having a common TUID system won't solve this challenge, but it will be a starting point.

Influence on Shipping Cost

The two phases impact shipping costs in very different ways. Most of the shipping expenses are typically incurred (i.e., checks are written) during the transport phase. However, those disbursements are typically decided and committed during the negotiation phase.

Negotiation Decisions	Transport Decisions
• **Choose Origin and Destination** • **Arrange Packaging** • **Set Route** • **Select Carrier** • **Schedule Shipment** • **Add Special Handling Requirements** • **Define Urgency**	• Equipment Selection • Driver Selection (maybe) • Maintenance Strategy

If a shipper wants to control their costs, the negotiation phase is the place to do it. In that phase, the shipper still has discretion to change schedules, reroute materials, find alternative carriers, and rethink their packaging. If the US trucking industry wants to streamline its operations, the negotiation phase would seem to be the place to do it.

Appendix 2

TUID Source Ingredients

The TUID is a hashed string built from specified text inputs. It is intended to be unique for all shipments on a global basis. It is also intended that a potential user can generate it unilaterally, independent from any central authority. The construction is as follows:

- **ctry** = text string with modified telephone country code, e.g.,
  - USA = "0"
  - Canada = "1"
  - …
  - Mexico = "55"
- **cno** = text string with nationally unique company id number … e.g.,
  - a federal tax EIN in USA, or a social security number in USA if Shipper is an individual unaffiliated with a company or organization
  - analogous company or individual IDs elsewhere (e.g., VAT number in Europe).
  - These must be consistent within a country of origin, but not necessarily across countries
- **cid** = text string id that is unique within the given company
  - A company is required to use a numbering or naming system that guarantees there will be no repetitions within the likely operational bounds of the **cno**.
  - There is a role for TMS, ERP, CRM or WMS to generate compliant ID automatically

Construction

With these components,

TUID = *hashfunction*(ctry+cno+cid)

The resulting TUID will be a long text string that is unique to a specific intended shipment to a specific Recipient (e.g., customer).

Example

Inputs:

ctry = "0" – US country code
cno = "12-3456789" – Typical US Employer Identification Number (EIN)
cid = "ABM123456789" – shipment identification code specific to shipper with above EIN

hash input = "0+12-3456789+ ABM123456789"

Possible Hash TUID Output

Potential hashed values for the example data with different hashing algorithms:

- MD5 Hash = 9e8ee3d6fc4602114c8f3808659f99d0
- SHA-1 Hash = f29c31c557505006d798cf74468189a858c1a69f
- SHA-256 Hash = bc587306fbdb5755c51a461a9d3801b402d9051a15d0389216536afa4a2e1467

Notes:

- *The guiding principle is to use **ctry** and **cno** to narrow the ID to a specific Shipper. Then the Shipper can add whatever internal counter or ID they find convenient to identify the shipment* *within that organization's logistics history. If the ctry and cno are unique, there is no possibility that one Shipper's TUID will ever match that of another Shipper.*

- *The intent of the hashing step is to create an ID that is effectively meaningless except as a label. If any party wants to associate other information with the TUID, they can bundle it with the TUID in a larger data package, but the TUID cannot be reverse engineered to reveal anything useful. The TUID is, by itself, physically incapable of leaking business information.*
- *Various hashing functions may generate shorter or longer hash strings. Finding the optimal encoding method is an avenue for research and refinement.*

Appendix 3

A Neutral, Transparent Governance System

As described above, the kernel of the TUID idea is simple and narrow.

- An identifier is created at the very earliest stage of the shipping process
- The Shipper shares it with the intended Recipient.
- (Optionally) the Shipper may submit it to a neutral history repository.

That is a potential platform for many industry innovations, especially in the processes where Shippers deal and negotiate with Brokers, DFMs and Carriers. However, as the TUID is used by more stakeholders for more purposes, those stakeholders will have concerns about security, fairness, and competitiveness. Managing these decisions surrounding the TUID and its uses requires a formal, transparent, democratic Governance System, where every interested stakeholder can have their say if they choose.

DFMDC has built a platform to govern exactly this type of data exchange system. The application was built to encourage the elimination of duplicate (aka Phantom) shipment booking data from the US trucking and logistics system. The intended stakeholders are participants in the Digital Freight Matching (DFM) industry and the proposed model requires the existence of a TUID such as the one described here. The proposed Governance structure is shown in Figure 5.

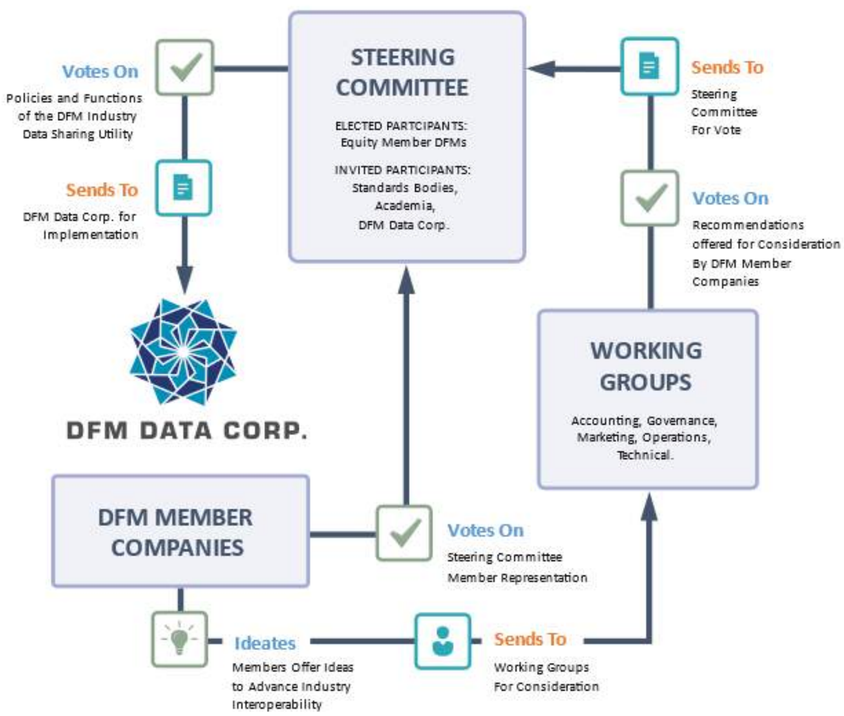


Figure 5 - DFM Data Corp Governance System

In this system, stakeholders are invited to become members of a consortium or group. The full body of stakeholder Members elects a smaller Steering Group which appoints and oversees the actions of specialized Working Groups. The Working Groups generate proposed changes which are sent to the Steering Committee for review and voted approval. Any accepted initiatives are then handed to the consortium staff with sufficient budget to implement. Once this model is up and running, all key decisions (including membership and usage fees) will be determined by the elected Steering Committee for the benefit of the Members.

Example Application - A Phantom Data Cleaning System

When Shippers, Brokers, DFMs and Carriers exchange information prior to booking a load, increasingly digitized systems allow the prospective shipments and trucking assets to become listed in many marketplaces. Often, listings will be copied from an origin to many other markets, possibly without the full knowledge of the shipper. When the load is eventually booked, many of these duplicate listings are not removed and remain in place, causing confusion and lost manhours.

DFMDC has pioneered a Governance System and a Governance Control Center software service to help organize an attack on this problem. However, this problem cannot fundamentally be solved unless a **Transport Unit ID (TUID) system is in place across a meaningful segment of the US trucking industry.** If the industry adopts a TUID, however, the cleaning process is pretty simple as shown in Figure 6.

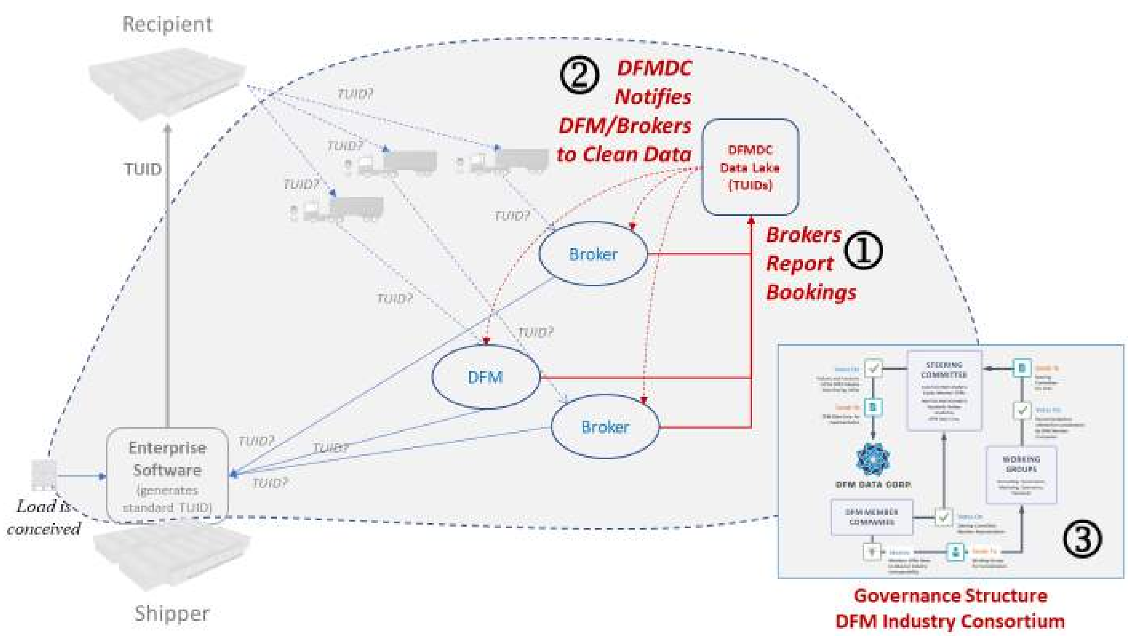


Figure 6 - Governance System for Removing Phantom Data

1. DFMs and Brokers are encouraged to report bookings to DFMDC's Data Lake. This booking notification includes the TUID.
2. DFMDC's Data Lake notifies participating DFMs and Brokers that a booking has occurred and shares the referenced TUID. Alternatively, DFMs and Brokers may elect to check their own listings against TUID bookings in the DFMDC Data Lake. Either or both methods are straightforward.
3. DFM Industry Consortium's Governance Structure manages:
    a. Recruiting and management of DFM and Broker industry members
    b. Maintenance and enhancement of the messaging infrastructure
    c. Consideration for the inclusion of additional types of data
    d. Reporting systems to share industry-wide data with Consortium members

e. Fees and services to members.

This is one of many scenarios where an industry standard TUID could empower major operational improvements in the US Trucking Industry.

