



# Intelligent Document Processing: Digital Examination of Freight Documents

Thought Leadership Paper

## Abstract

The examination of freight documents is a vital process in international trade, ensuring legal compliance, facilitating financial transactions, and enhancing supply chain efficiency. Key documents, such as bills of lading, commercial invoices and certificates of origin, play a central role. Intelligent document processing is able to significantly improve this process, even including challenges like identifying document discrepancies and overcoming language issues. As global trade continues to evolve, the examination of freight documents remains a cornerstone of efficient, secure, and compliant logistics operations, contributing to the smooth flow of goods across borders and the growth of international commerce.

Version:	1.0
Date:	September 2023
Author:	Peter Ortmanns
Email:	Peter.ortmanns@iriscorporate.com
Website:	www.iriscorporate.com

## Table of Contents

Abstract.....	1
Executive Summary .....	3
1. Intelligent Document Processing .....	4
2. Freight Document Processing .....	6
2.1 Importance of Freight Documents .....	6
2.2 Freight documents in general.....	6
2.3 Key Freight Documents in detail.....	7
2.4 Documents and Data .....	8
3. Documentation in the Shipping Process .....	9
3.1 Incoterms.....	9
3.2 A sample process .....	11
4. Intelligent Document Processing for Freight Records .....	13
5. AI, ML, OCR, IDP: How does it work? .....	15
5.1 Document Classification .....	15
5.2 Data Extraction .....	16
5.3 Main features of IRISXtract™ .....	16
6. Conclusions.....	18

## Table of Figures

Fig. 1: Automated Document Processing in short .....	4
Fig. 2: Application Scenarios .....	5
Fig. 3: Documents & Data fields .....	8
Fig. 4: Incoterms explained by Freightos .....	10
Fig. 5: The bill of lading in the shipment processes .....	11
Fig. 6: A typical software architecture .....	14

## Executive Summary

The examination of freight documents is an essential part of the logistics process for transporting goods, whether domestically or internationally. These documents play a crucial role in facilitating the movement of cargo while ensuring compliance with customs regulations and international trade laws. They are vital for customs clearance, security and safety, trade regulations, ownership and control, risk management, logistics and tracking, as well as payment and settlement. There are several documents involved in the shipping process. Here is a selection of the main documents:

- Commercial Invoice: Contains details about the goods being shipped.
- Certificate of Origin: Indicates the country where the goods originated.
- Bill of Lading: Serves as a contract between the shipper and carrier.
- Packing List: Lists the contents and packaging details of the shipment.
- Letter of Credit: Ensures payment to the exporter.

These documents help to ensure that goods are transported smoothly from point A to point B while complying with legal and trade requirements. The examination process involves verifying that the cargo complies with laws and regulations when entering a country. It helps identify potentially hazardous or illegal items, ensures security and safety in international trade, and demonstrates compliance with trade regulations, restrictions and commercial agreements between the consigner and the consignee.

However, the manual processing of these documents is inefficient, related to high costs and error-prone what results in a lack of data accessibility and accuracy, long cycle times with difficulties to keep SLAs, disgruntled customers, suppliers, and even co-workers and prevents from scaling the business easily. Hence, the main objectives and benefits of introducing computer aided process automation are:

- Costs: eliminate redundancies, automate simple/repetitive tasks.
- Time: faster handling time, lower holding time, better SLA fulfillment for your customers.
- Quality: rule sets to avoid errors and to increase accuracy.
- Transparency: Enhance controlling, data analytics, compliance, and auditability.
- Flexibility: standardized process can be adapted to changing requirements.
- Business Development: resources freed for valuable work or to scale the number of cases.

To achieve such objectives document understanding technologies supporting the automated processing must be capable to

- separate files to be processed that contain different documents into single files,
- classify these documents by identifying their type,
- recognize type specific data, extract it independent from the positioning of the data,
- validate the accuracy of data and simultaneously consolidate the data by document and case intrinsic crosschecks as well as matching with data from other sources and
- finally normalize the data format for a seamless integration into subsequent business software for further processing.

## 1. Intelligent Document Processing

Since 35 years I.R.I.S. - A Canon Company helps more than 40 million users worldwide to transform their processes into digital successfully. Our Intelligent Document Processing platform IRISXtract™ for Documents imports documents, classifies them and captures data automatically. Data will be consolidated and normalized. Thus, unstructured data is transformed into structured information that seamlessly drives the rest of the process landscape.

Intelligent Document Processing (IDP) is a technology-driven approach that combines artificial intelligence (AI) and machine learning (ML) techniques with automation to extract, interpret, and manage data from unstructured documents. IDP systems are designed to streamline and improve the efficiency of document-centric processes within organizations.

Fig. 1: Automated Document Processing in short



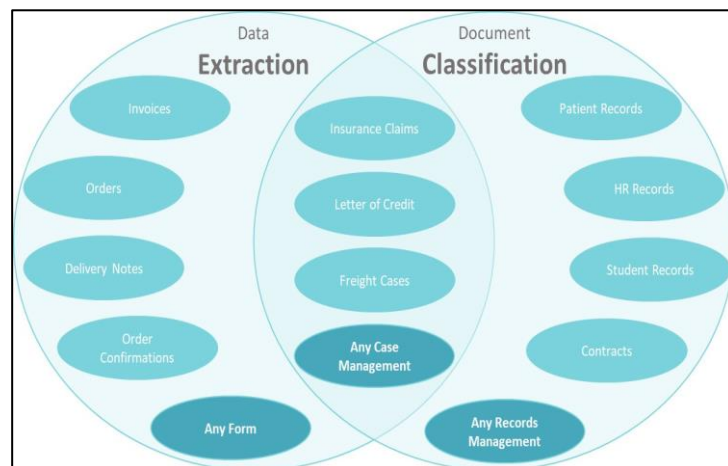
IDP solutions begin by ingesting a wide variety of unstructured documents, including text documents, images, scanned files, PDFs, emails, and more. These documents can contain structured and unstructured data. Usually the IDP software starts to use methods for the optimization of the image in order to in-

crease the accuracy of Optical or Intelligent Character Recognition (OCR: machine written text; ICR: handwritten text). Once the documents are captured, IDP systems use AI and ML algorithms to automatically separate different documents that are included in one file or to distinguish between the main document and attachments to that document. IDP can also automatically classify documents into predefined categories. For example, it can distinguish between invoices, purchase orders, contracts, receipts, and other document types that are known by the software based upon machine learning. Knowing the class of the document IDP software extract relevant data from them. This includes extracting specific information, such as names, dates, addresses, document numbers, line items in tables and more, from within the documents. Whilst doing so IDP systems can validate and verify extracted data against existing databases, business rules, or reference data (crosschecks between the documents) to ensure accuracy and consistency. Extracted data is often transformed into structured formats, such as database records, spreadsheets, or other digital formats, making it easier to integrate such normalized data into downstream systems. Further there are compression technologies in use to minimize the size of an exported pdf file. If additionally Natural Language Processing (NLP) capabilities are embedded in IDP then sentiment analysis, language translation, and context understanding, are enabled and allow more advanced processing of text-based documents.

Thus, adapted for particular applications such technologies allow to configure rather any kind of automated document processing scenario starting with simple mailroom routing by identifying for instance the related department up to more particular classification and data extraction requirements. Such are for example department or process specific records management processes like:

- HR Records
- Legal Cases
- Procurement Case
- Patient Records
- Patent Records
- Student Records
- Governmental Cases
- (Mortgage) Loan Processing
- Customer Onboarding
- Real Estate Records
- Customer or Supplier Records
- Letter of Credit
- Freight Records

Fig. 2: Application Scenarios



This paper is focussing on the use of intelligent document processing for freight documents since the examination of freight documents is not without its challenges. Manual processing can lead to delays, disputes, and financial losses. Some common challenges include:

- **Document Discrepancies:**  
Discrepancies between documents, such as inconsistent quantities or descriptions, can lead to delays in customs clearance and disputes between parties.
- **Language and Translation:**  
In international trade, documents are often prepared in multiple languages. Accurate processing independent from the language is crucial to avoid misunderstandings and errors.
- **Manual Processes:**  
Many aspects of document examination are still reliant on manual processes, which are time-consuming and prone to human error.
- **Regulatory Changes:**  
Frequent changes in import/export regulations and trade agreements can create confusion and require constant updates to documentation practices.
- **Fraud and Forgery:**  
The complexity of international trade provides opportunities for fraudulent activities, such as document forgery, trade-based money laundering and misrepresentation of goods.

The next chapter will dive into more details of these challenges by explaining freight document processing by elaborating the requirements for intelligent document processing.

## 2. Freight Document Processing

The processing of freight documents is a crucial aspect of international trade and logistics. It involves the scrutiny, verification, and processing of various documents associated with the transportation of goods from one location to another. In a globalized world where goods are constantly moving across borders, the accuracy and integrity of these documents are paramount to ensure the smooth flow of goods and compliance with international regulations. This chapter delves into the examination of freight documents, discussing its importance, the key documents involved, the examination process, challenges, and technological advancements that have reshaped this field.

### 2.1 Importance of Freight Documents

Freight documents serve as the backbone of international trade, providing a paper trail that facilitates the movement of goods and ensures the rights and obligations of all parties involved are clearly defined. These documents are not just pieces of paper; they represent the legal, financial, and logistical aspects of a transaction. Several factors underscore the importance of freight documents:

- **Legal Compliance:**  
International trade is subject to a complex web of laws and regulations. Freight documents, such as bills of lading and certificates of origin, help shippers and consignees ensure that their transactions adhere to these regulations. Failure to comply can result in legal consequences and delays in the movement of goods.
- **Financial Transactions:**  
Freight documents are often linked to payment mechanisms, such as letters of credit. Banks use these documents to release funds to the exporter only when the conditions specified in the document are met. Accurate examination is crucial to prevent disputes and financial losses.
- **Risk Management:**  
Examination of freight documents is an essential part of risk management in logistics. By verifying the accuracy of documents, logistics professionals can identify potential issues early, such as discrepancies in quantity or quality, and take appropriate actions to mitigate risks.
- **Supply Chain Efficiency:**  
Accurate and timely examination of freight documents contributes to the efficiency of the supply chain. Delays caused by document discrepancies can disrupt the flow of goods, increase costs, and erode customer trust.

### 2.2 Freight documents in general

In a freight case, there are various types of documents that may be involved, depending on the nature of the shipment, the mode of transportation, and the regulatory requirements of the countries involved. The following lists the common document types that can be associated with a freight case:

- Bill of Lading
- Commercial Invoice
- Packing List
- Certificate of Origin
- Proforma Invoice
- Certificate of Inspection
- Customs Declaration
- Export License
- Import License
- booking confirmations
- shipping instructions
- cargo tracking information
- Dangerous Goods Declaration
- Certificate of Insurance
- Importer Security Filing
- Air Waybill / Sea Waybill
- Export Packing List
- Certificate of Conformity
- Bill of Entry
- Letter of Credit
- Inspection Certificates

### 2.3 Key Freight Documents in detail

The examination of freight documents involves a range of documents, each with a specific purpose in the transportation process. Some of the key documents include:

- **Bill of Lading (B/L):**  
The bill of lading is perhaps the most important freight document. It serves as a receipt for the goods, evidence of the contract of carriage, and a document of title. It is used to prove ownership of the goods and facilitate their transfer between parties.
- **Commercial Invoice:**  
The commercial invoice details the specifics of the goods being shipped, including their description, quantity, price, and terms of sale. It is essential for customs clearance, calculating duties and taxes, and facilitating payment between the buyer and seller.
- **Packing List:**  
This document provides a detailed list of the contents of a shipment, including information on how the goods are packed. It aids in customs clearance, helps verify the accuracy of the shipment, and assists in determining the appropriate handling and storage requirements.
- **Certificate of Origin:**  
A certificate of origin is used to establish the country of origin of the goods. It is essential for complying with trade agreements and determining applicable tariffs and import restrictions.
- **Letter of Credit (if applicable):**  
A letter of credit is a usual payment method and ensures that the terms and conditions specified in the letter of credit match the information on the freight documents, including the bill of lading and commercial invoice.

The purpose of an automated examination of freight documents is to streamline and enhance the efficiency, accuracy, and security of the document verification process in international trade and logistics. Automated examination employs various technologies, including AI, ML, OCR, and electronic document management systems.

In summary, the purpose of automated examination of freight documents is to modernize and optimize the document verification process in international trade and logistics. By leveraging technology to efficiently and accurately handle documents, businesses and regulatory authorities can ensure smoother, more compliant, and cost-effective international trade operations.

## 2.4 Documents and Data

To approve a case in the context of international trade and logistics, several types of data need to be extracted and validated from various freight documents. The specific data requirements may vary depending on the nature of the case and the regulatory requirements of the countries involved. However,

Fig. 3: Documents & Data fields

Class \ Datafield	Commercial Invoice	Bill of Lading <sup>2</sup>	Packing List	Certificate of Origin	Letter of Credit	Insurance	Sea/Airway Bill <sup>2</sup>
Class	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Document Date	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Departure Date		<input checked="" type="checkbox"/>					
Arrival Date		<input checked="" type="checkbox"/>					
Issuing Country				<input checked="" type="checkbox"/>			
LC Number	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Invoice Number	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
B/L Number		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
P/L Number			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
Reference Number				<input checked="" type="checkbox"/>			
Document Number							<input checked="" type="checkbox"/>
Loading Port	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>
Loading Country	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>
Discharging Port	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>
Discharging Country	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>
Vessel/Carrier	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Voyage Number		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>
Incoterms	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>			
Exporter	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Importer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Notify Party		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>
Advising Bank					<input checked="" type="checkbox"/>		
Issuing Bank					<input checked="" type="checkbox"/>		
Goods	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
Description	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
Quantity	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
Unit Price	<input checked="" type="checkbox"/>						
Totals	<input checked="" type="checkbox"/>						
Weight		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
Measurement		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
Markings			<input checked="" type="checkbox"/>				
Container Number(s)		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
Total Amount	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/> <sup>1</sup>	<input checked="" type="checkbox"/>	
Currency	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Payment Terms	<input checked="" type="checkbox"/>						
Credit Terms	<input checked="" type="checkbox"/>						

here are some common data elements that typically need to be extracted and validated from key freight documents.

Fig. 3 lists the data fields per document class of the main documents of a case. These data usually need to be extracted from the document, validated, and normalized. If a field isn't a specific field such as the class or the document date, then the consolidation of data across the documents is due. Any discrepancies or inconsistencies should be addressed and resolved to ensure that the case can be approved for further processing, including customs clearance and payment release. Depending on the specific case and the parties involved, additional documents and data elements may also be required for validation.

<sup>1</sup> In case of an insurance the insured amount must be the invoice amount plus 10%.

<sup>2</sup> A bill of lading and an air/sea waybill are exclusive documents that cannot be part of the same case. They are both used for different purposes in international shipping. A bill of lading is a negotiable document that can be transferred from one party to another, while the air/sea waybill is a non-negotiable document. A bill of lading requires the presentation of the original document at the destination port to release the goods, while an air/sea waybill does not require any document presentation. A bill of lading is usually used when there is a credit transaction or a third party involvement in the trade, while an air/sea waybill is usually used when there is a cash transaction or a direct trade between the buyer and seller.



### 3. Documentation in the Shipping Process

Processing freight documents is a meticulous process that involves various stakeholders, including shippers, carriers, customs authorities, ports, and in case of letter of credit even banks. The process can be divided into some key steps:

- **Document Submission:**  
When a shipment is initiated, the shipper or exporter prepares the necessary documents and submits them to the relevant parties, such as the carrier, the buyer, and customs authorities.
- **Verification of Documents:**  
The first step in the examination process is to verify the accuracy and completeness of the submitted documents. This includes checking for discrepancies between the bill of lading, commercial invoice, and other documents.
- **Customs Clearance:**  
Customs authorities play a vital role in the examination process. They review the documents to ensure compliance with import/export regulations, calculate duties and taxes, and decide whether to release the goods for onward transportation.
- **Letter of Credit Examination:**  
In cases where payment is facilitated through a letter of credit, banks carefully examine the documents to ensure they match the terms and conditions specified in the letter of credit. Any discrepancies can lead to delays in payment.
- **Communication and Resolution:**  
If discrepancies or issues are identified during the examination process, communication between the relevant parties is essential. This may involve negotiations, corrections to documents, or additional documentation to resolve the issues.
- **Release and Transportation:**  
Once all discrepancies are resolved, and the documents are in order, the goods are released for transportation to their final destination. This includes the physical movement of the goods and the necessary notifications to the consignee.

#### 3.1 Incoterms

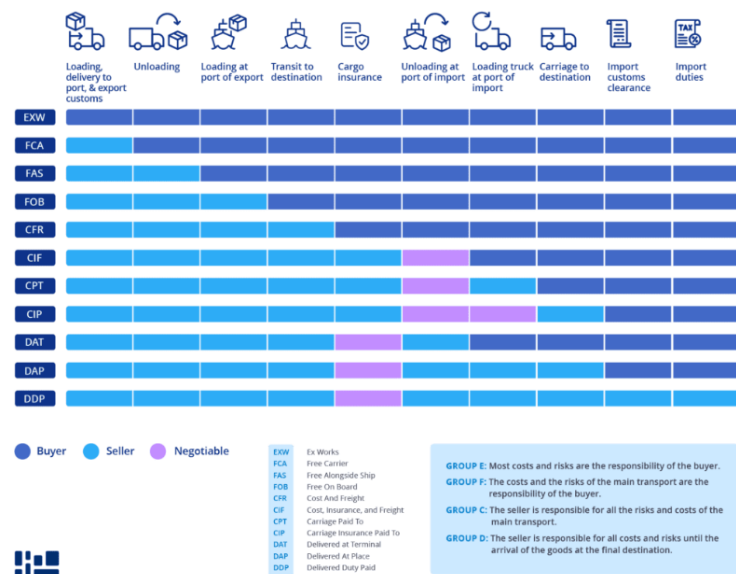
Incoterms (International Commercial Terms) have a significant impact on the processing of freight documents in international trade. Incoterms are a set of standardized trade terms published by the International Chamber of Commerce (ICC) that define the rights and responsibilities of buyers and sellers regarding the delivery of goods. These terms specify when risk and ownership of the goods transfer from the seller to the buyer and also indicate the point at which the seller's obligations for transportation and documentation are fulfilled. Here's how Incoterms influence the examination of freight documents:

- **Determination of Responsibility:**  
Incoterms define whether the seller or the buyer is responsible for obtaining specific documents and complying with customs and export/import regulations. For example, under the term "Ex Works" (EXW), the seller's responsibility for document preparation is minimal, while

- under "Delivered Duty Paid" (DDP), the seller must provide a wide range of documents and handle customs clearance.
- Choice of Incoterms Affects Document Requirements:  
The choice of Incoterms in a sales contract directly influences the specific freight documents that need to be examined and provided by each party. For example:
  - Under "Free On Board" (FOB), the seller is responsible for the bill of lading, while the buyer may be responsible for additional documents such as import licenses.
  - Under "Cost and Freight" (CFR), the seller is responsible for the bill of lading, but the buyer typically handles additional documents, including insurance.
  - Under "Delivered at Place" (DAP), the seller is responsible for delivery and unloading, but the buyer may be responsible for import customs clearance and payment of import duties.

- Transfer of Risk and Ownership:  
Incoterms establish the point at which risk and ownership of the goods pass from the seller to the buyer. The examination of freight documents plays a role in verifying this transfer and ensuring that the documents accurately reflect the agreed-upon Incoterms. For instance, under "CIF" (Cost, Insurance, and Freight), risk transfers from the seller to the buyer when the goods are loaded onto the vessel, which is reflected in the bill of lading and insurance certificate.

Fig. 4: Incoterms explained by Freightos



<https://www.freightos.com/freight-resources/incoterms-plain-english-freight-shipping-guide/> (2023-09-28)

- Impact on Payment and Financing:  
Incoterms can also influence the payment terms and financing arrangements between the buyer and seller. The examination of documents such as letters of credit is crucial in these scenarios to ensure compliance with the agreed-upon Incoterms and to facilitate the payment process.
- Customs and Regulatory Compliance:  
The choice of Incoterms can affect the customs clearance process, as different terms may require different documentation and procedures. Shippers and buyers must ensure that the freight documents align with the chosen Incoterms to avoid customs issues and delays.

- **Documentation Accuracy:**  
Given that Incoterms define the responsibilities of each party, it is essential that freight documents accurately reflect the agreed-upon terms. Any discrepancies in the documents can lead to disputes and potentially impact the release of goods at the destination.

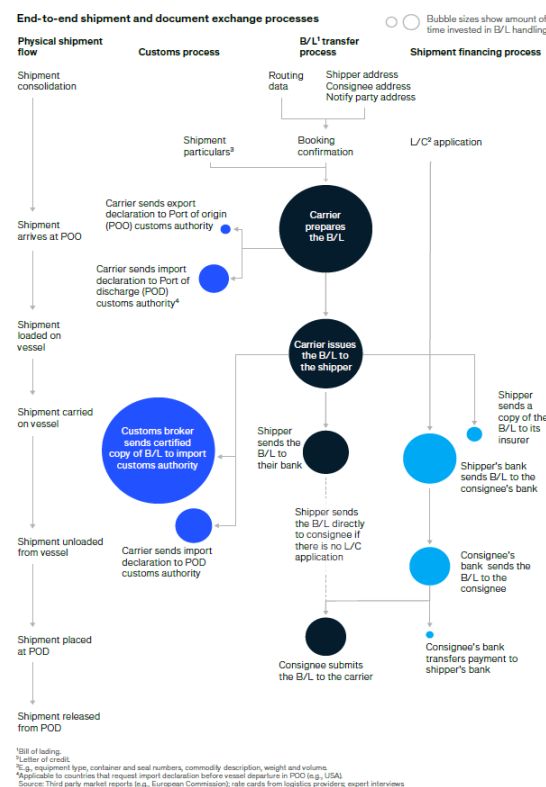
In summary, Incoterms are a crucial aspect of international trade contracts, and they play a pivotal role in determining the obligations and responsibilities of both buyers and sellers. These terms directly impact the examination of freight documents by dictating which documents are required, who is responsible for their preparation, and at what point risk and ownership transfer. Therefore, a clear understanding of Incoterms is essential for shippers, importers, and exporters to ensure smooth and compliant international trade transactions.

### 3.2 A sample process

In a summary the process of examining and analyzing shipment documents depends on lots of variables such as the documents effectively involved in a particular case, the Incoterms, the stakeholder

Fig. 5: The bill of lading in the shipment processes

The bill of lading is the center of many shipment processes.



D. Casanova, D. Dierker, L. Hausmann, B. Jensen, and J. Stoffels: The multi-billion-dollar paper jam: Unlocking trade by digitalizing documentation, McKinsey & Company, 2022, p. 4

of the goods locally. An agent contacts the supplier locally to arrange the movement of the freight.

who needs to process the documents for his specific purposes, the specific software architecture to which the data needs to be submitted, and also how the deal is financed. Figure 5 by McKinsey illustrates the process from the point of view of the bill of lading in a trade case that is financed by letter of credit.

The shipping process from beginning to end in case of EXW terms and engaging an independent forwarder and using **letter of credit** as payment method starts with the importer ordering the goods from a supplier who then provides a **pro-forma invoice** which is like a quote. Based upon this a Purchase Order is created, and the buyer hires a freight forwarder to handle the transport of the goods. The freight forwarder confirms the **Incoterms** with the buyer and the supplier. The buyer obtains a **letter of credit** from the bank to pay the goods. The supplier's bank approves the letter of credit so that the manufacturing of the goods can start. The supplier provides the buyer an **order confirmation** and the **commercial invoice** and the freight forwarder contacts his overseas partner to arrange the movement

The supplier then provides **all the documents** necessary for the export, which will be handled on the buyer's behalf by the freight forwarder. With the documentation in order, the supplier books the shipment for export. When a carrier arrives to pick up the goods at the supplier's warehouse a **bill of lading** is issued. The buyer will be required to present the **bill of lading** in order to secure the release of the shipment and to claim the ownership over the goods. The freight forwarder may manage this process. The goods must be customs cleared and delivered to the agreed location of the buyer and the supplier completes the export declaration. Only then the goods are placed in international transit so that the goods will arrive in buyer's country for import clearance. After import clearing the goods are finally delivered from the destination country's port to the buyer. With the documentation at hand the seller's bank then can claim the payment on behalf of the seller from the buyer's bank.<sup>1</sup>

---

<sup>1</sup> If you want to learn more about the specific requirements on LC processing please watch the recording of our presentation on the FSI Conference 2021 ["Intelligent Letter of Credit processing"](#).

#### 4. Intelligent Document Processing for Freight Records

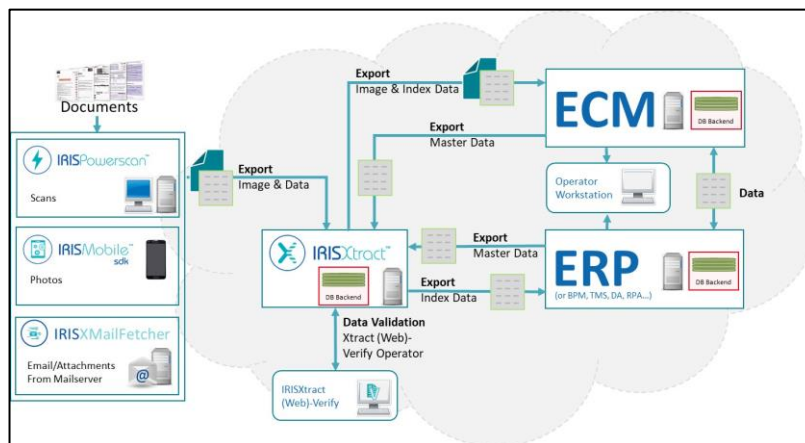
Since IDP is a technology that combines AI, ML, and OCR to automate the extraction, validation, and processing of information from various types of documents, IDP can significantly enhance the examination of freight documents in several ways:

- **Document Separation and Classification:**  
IDP can automatically separate documents contained in one source file and classify freight documents based on their type, such as distinguishing between bills of lading and packing lists. This classification helps route documents to the appropriate workflows for further processing.
- **Automated Data Extraction:**  
IDP systems can automatically extract critical data from freight documents such as bills of lading, commercial invoices, packing lists, and certificates of origin. This includes information like product descriptions, quantities, prices, dates, and vendor or consignee details.
- **Data Validation and Verification:**  
IDP can validate the extracted data against predefined rules and patterns. It can cross-reference data across multiple documents to ensure consistency and accuracy, reducing the risk of human errors in document examination.
- **Exception Handling:**  
When discrepancies or anomalies are detected in the extracted data, IDP can flag these as exceptions. These exceptions can be reviewed by human operators for resolution, streamlining the examination process.
- **OCR Accuracy Improvement:**  
IDP often incorporates advanced OCR technology to accurately capture data from scanned or handwritten documents, even when the quality of the documents is suboptimal. This ensures high accuracy in data extraction.
- **Data Enrichment:**  
IDP systems can augment extracted data by cross-referencing it with external sources or databases to provide additional context and information about the goods or trading partners mentioned in the documents.
- **Customizable Workflows:**  
IDP solutions can be configured to support specific workflows and business rules. This enables organizations to tailor the document examination process to their unique requirements and compliance needs.
- **Integration with Other Systems:**  
IDP can integrate seamlessly with other supply chain and logistics systems, such as transportation management systems (TMS) or enterprise resource planning (ERP) systems. This ensures that the data extracted from freight documents can be easily shared and used across the organization.
- **Scalability and Efficiency:**  
IDP can handle large volumes of documents quickly and efficiently, allowing organizations to

scale their document examination processes as their business grows. This efficiency reduces processing times and labor costs.

- **Audit Trails and Compliance:**  
IDP maintains detailed audit trails of document processing activities, ensuring transparency and compliance with regulatory requirements. These audit trails can be invaluable during audits or investigations.
- **Real-time Monitoring and Alerts:**  
Organizations can monitor the document processing in real-time and receive alerts for critical issues or exceptions, allowing for proactive intervention when necessary.
- **Reduced Manual Work:**  
By automating data extraction and validation, IDP reduces the need for manual data entry and repetitive, time-consuming tasks, freeing up employees to focus on more strategic and value-added activities.

Fig. 6: A typical software architecture



(IRIS Mobile Capture SDK) and email fetching (IRIS Mailfetcher).

Figure 6 illustrates how IRIS IDP technology is integrated in an architecture of business software suites. The core software here is IRIS Xtract™. This IDP platform is equipped with all necessary technology to fulfill the up-front listed objectives. You will learn more about it in the next chapter.

If needed IRIS also provides software components to support scanning (IRIS Powerscan), mobile capturing

## 5. AI, ML, OCR, IDP: How does it work?

The exact scope of AI is shifting with progress made in the field; for example, optical character recognition (OCR) is nowadays often not included in AI, as it has become a routine technology. On the other hand, systems that are using complex algorithms (e.g. advanced Regular Expressions) may appear intelligent and are thus often included in AI, although self-learning techniques isn't a feature of this approach. AI is applying self-learning techniques.

ML is a sub field of AI. ML includes algorithms that learn inherent patterns from a given data representation and is able to create predictions or decisions concerning these data samples. Deep Learning is a sub field of Machine Learning, concerned with learning data representations by using techniques based on information processing in an Artificial Neural Network (ANN). Each "neuron" in such a network has very basic functionality. Often just taking an input, transforming it, then passing it on.

The network's power is achieved by combining many neurons in a layer and then connecting these layers to each other. An input layer receives data samples, an output layer presents a prediction about the given sample. In between these layers are so-called hidden layers, which hold weights and that are able to create new representations of the data. If an ANN has multiple hidden layers, it is called a Deep Neural Network (DNN). A special type of a DNN is the Convolutional Neural Network (CNN). Our Optical Character Recognition engine named High-Quality OCR (HQ-OCR) is based on such a CNN. Unlike traditional OCR engines, deep learning automates feature learning from raw pixel data through a combination of Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) layers. Thanks to this advanced AI technology, HQ-OCR excels at handling font, size, and orientation variations, enabled by its capability to learn complex features. These engines can read entire words and lines, resolving ambiguities on the fly and grants exceptional performance on low-quality document images.

### 5.1 Document Classification

Learning algorithms are mostly either supervised or unsupervised. Supervised means the training of the system is performed by using labeled data. By doing so the system can calculate and adapt to the difference between the true label and label predicted by the system.

Logistic Regression is a technique just as much based in statistics as in ML. In the case of document classification, IRISXtract™ looks at the words in each given document as features and defines a metric of relevance for each feature. Then, IRISXtract™ introduces a coefficient for each feature that has to be "learned", meaning iteratively improved by feeding labeled samples into the algorithm (supervised learning) and improving the prediction. The core components inside IRISXtract™ enabling this are XClassify and XContext. For the learning part, XClassify uses Stochastic Gradient Descent to minimize the error function, which can be seen as a distance function between prediction and truth. After learning the weight matrix of the algorithm, it can use these weights to make predictions about the class of an unknown document. Such a prediction comes in the form of a confidence value (0-100%) for each class, stating how confident the algorithm is that the document belongs to that class.

## 5.2 Data Extraction

The technology and approach for information extraction have undergone significant evolution, progressing from the initial versions of Optical Character Recognition (OCR), data capture for forms using layout templates, and eventually to the introduction of freeform capture technologies. The latter represents an intelligent and adaptive capture strategy, involving semantic analysis of textual documents to identify predetermined keywords and patterns.

These technologies can also recognize table structures, perform fuzzy logic data matching, and consolidate information against master data. Such master data can be provided by data bases of other software or can be generated inside IRISXtract™. Such internal data allows to compare extraction results of data from different documents. Notably, such freeform approaches eliminate the need for fixed layout templates and are thus referred to as **Intelligent Document Recognition (IDR)** or **Intelligent Document Processing (IDP)**. Instead of relying on rigid geometric templates, a semantic document model defines fields, keywords, and business logic for specific document categories.

The XContext engine implements such a freeform strategy and significantly reduces configuration efforts and hence total cost of ownership for customers. Since decades, XContext has demonstrated its success in processing semi-structured documents as integral component of the IRISXtract™ platform and is even capable to address unstructured digital mailroom scenarios.

## 5.3 Main features of IRISXtract™

- **Multi-Channel Import:**  
Import of scanned paper documents, electronic documents from emails, portals or mobile devices and that in different file formats.
- **Multilingual:**  
The IRIS HQ-OCR text recognition can recognize more than 140 languages, using CNN even on documents with poor image resolution.
- **Hybrid formats:**  
Documents that include multiple languages at the same time are processed, as well as documents that include handwriting and/or typewriting.
- **Document separation:**  
The software recognizes the beginning and end of a document and can thus separate all documents contained in a file, for example a PDF file.
- **Document classification:**  
The software automatically distinguishes document types. This is essential for case management, because the further process depends on
  - whether a certain document type is relevant for further processing or only needs to be stored indexed in order to be found if necessary or
  - whether a process is complete, i.e. all document types required for further processing have been provided and
  - to which next process or to which team a document or case must be transferred.
- **Data capture:**  
The software automatically extracts data by the help of different methods and technologies



also including supervised trainings by the user. Data fields are, for example, case or document numbers, senders and recipients, dates and deadlines, amounts, tables and lists, or general process parameters and conditions.

- **Validation, consolidation and normalization:**  
The data must usually comply with certain formats or structures, be logically or mathematically correct, match third-party data, some may be missing, others not, so that, depending on the result, further processing can be fully automated or a separate exception handling process is triggered.
- **Multi-Application:**  
IRISXtract™ offers parallel operation of different solutions such as:
  - Dedicated documents (offers, delivery bills, orders and invoices)
  - Case management in human resources, patent departments, legal departments, law firms, clinics, insurance companies, banks, citizen services or
  - Process cascades, in which documents are first sorted by class or department, and then the desired indexes are extracted depending on this.
- **Multi-Tenancy:**  
IRISXtract™ offers parallel operation of a certain solution in different configurations to meet decentralized needs for example at different locations or for different purposes.
- **Business Intelligence:**  
Process data from the use of the software is stored and prepared for process control, auditability, tracking and tracing, and for a continuous improvement process.
- **Conversion and compression:**  
Data and documents are transferred to downstream applications in the required formats and structures seamlessly and in a transaction-safe manner. PDF Documents can also be compressed many times by using IRIS proprietary technology called iHQC (intelligent high quality compression) to save storage space and reduce transaction times.

In summary, IRISXtract is an Intelligent Document Processing platform that offers significant support to the examination of freight documents by automating and enhancing data extraction, validation, and processing. It improves accuracy, reduces processing times, and helps organizations ensure compliance with trade regulations, ultimately contributing to more efficient and error-free supply chain and logistics operations. IRISXtract™ is powered by proprietary technologies like OCR covering over 140 different languages and PDF compression. IRISXtract™ combines the most advanced technologies, including AI-driven text and index recognition, ML-based document classification and, in terms of hyper-automation, context-based learning of the software during data extraction by simultaneously checking data integrity and consistency.



Learn more about IRISXtract™ at work in logistic companies by watching the replay of our webinar [“Sailing the digital wave: Revolutionizing Freight Document Processing.”](#)

## 6. Conclusions

The examination of freight documents is a critical component of international trade and logistics. It ensures legal compliance, facilitates financial transactions, manages risk, and enhances supply chain efficiency. Key documents, such as bills of lading, commercial invoices, and certificates of origin, play a pivotal role in this process. However, the examination process is not without its challenges, including document discrepancies, language issues, and regulatory changes. Intelligent Document Processing with IRISXtract™ grants:

- **Efficiency:** Automation reduces the reliance on manual, time-consuming tasks, such as data entry and document cross-referencing. It allows for the rapid processing of a large volume of documents, resulting in faster document clearance and goods release.
- **Accuracy:** Automation significantly reduces the risk of human error in document examination. AI and OCR technologies can extract and compare data from documents with high precision, minimizing the likelihood of discrepancies and data entry mistakes.
- **Consistency:** Automated systems apply consistent rules and criteria when examining documents, ensuring that each document is assessed using the same standards. This consistency reduces the potential for bias or subjective judgment.
- **Real-time Data Access:** Automation enables real-time access to data, allowing customs authorities, shipping companies, and other stakeholders to quickly verify information and make informed decisions. This is especially valuable in time-sensitive situations.
- **Compliance:** Automation can help ensure compliance with complex international trade regulations, including customs requirements and trade agreements. It can flag discrepancies or issues that may lead to non-compliance.
- **Cost Reduction:** By reducing manual labor and improving efficiency, automated document examination can lead to cost savings for businesses and government agencies involved in international trade.
- **Risk Management:** Automated systems can analyze data and flag potential risks or irregularities, such as discrepancies in document details, suspicious patterns, or inconsistencies with trade sanctions lists. This helps in proactively managing risk.
- **Scalability:** Automated examination systems can easily scale to accommodate increased document volumes as trade volumes grow, making them adaptable to changing business needs.
- **Integration:** Automated systems can integrate with other supply chain and logistics technologies, such as transportation management systems (TMS) and warehouse management systems (WMS), for seamless data flow and enhanced supply chain visibility.
- **Customer Satisfaction:** Faster document processing and reduced delays in customs clearance lead to improved customer satisfaction and trust in the supply chain.

In conclusion, the examination of freight documents is not just a routine task but a critical function that underpins the entire international trade ecosystem. Its proper execution ensures the smooth movement of goods, compliance with regulations, and the integrity of financial transactions, making it indispensable in the modern world of global commerce.