Technical considerations for automatic discovery on the Peppol network

*Jelte Jansen*

*Version 1.0, 2021-05-18*

Table of contents

[1 Introduction 3](#_Toc69417052)

[2 The concept of automatic discovery 3](#_Toc69417053)

[3 Sending documents 4](#_Toc69417054)

[3.1 Discovery 4](#_Toc69417055)

[3.2 Data points for discovery 4](#_Toc69417056)

[3.3 Data enrichment from external sources 5](#_Toc69417057)

[3.4 Generation of Peppol BIS documents 5](#_Toc69417058)

[3.5 When discovery is not possible 5](#_Toc69417059)

[4 Receiving documents 6](#_Toc69417060)

[4.1 Publishing recipients 6](#_Toc69417061)

[4.2 Receiving and processing documents 6](#_Toc69417062)

[4.3 Routing based on recipient identifier 6](#_Toc69417063)

[5 Conclusion 7](#_Toc69417064)

# Introduction

This document describes a number of technical considerations for sending and receiving documents on the Peppol network, focused on interoperability regarding the automatic discovery of recipients.

Automatic discovery of recipients is of importance in a Peppol-first approach: If a document can be sent with Peppol, it should be, and the actual sender of the invoice should not have to manually configure this for each potential recipient, nor have bilateral agreements with them. Such an approach further expands the advantages of the Peppol network and the implementation and automation of e-invoicing.

That said, this document does \*not\* discuss Peppol-first itself, nor related approaches to expand the use of e-invoicing and the Peppol network.

The scope of this document is limited to the technical considerations regarding automatic discovery, such as the minimal requirements for automatic discovery to be possible, and potential pitfalls of certain implementation choices by the various parties in the 4-corner model.

This document focuses on invoices, but its principle should be applicable to other document types as well.

# The concept of automatic discovery

Whenever a supplier wishes to send an invoice to a customer, they may or may not be aware that the customer is reachable on the Peppol network; In a 'manual' scenario, it is bilaterally agreed that invoices are exchanged via Peppol, and a number of additional details are exchanged. One such detail is the identifier with which the receiver of documents can be reached on the network, such as their Chamber of Commerce or VAT number.

Often, additional details are agreed upon as well, such as order reference numbers, but also some details about the sender, with which the receiver may recognize the invoice as one that is from an existing business contact, which is then stored in their ERP system, and eases the process of acknowledging incoming invoices.

Such bilateral agreements also imply that the sender of documents perform additional steps when sending a document, which might reduce the effectiveness of automation, or introduce additional errors. For example, the sender might have to manually tell their own system that the agreed-upon method of transport is the Peppol network, and with which identifier.

Information of that type could be derived from data which is already present in the sender's system, the invoice itself, and the Peppol network. This is where automatic discovery becomes useful: The sender's system could automatically use the Peppol network if it is aware the recipient is present on the Peppol network, and fall back to other methods of exchange it the recipient is not.

While there will always be some information that needs to be agreed upon, such as order numbers, automation and reliability could be improved if the amount of that information is kept to a minimum.

However, there are a number of potential issues that could arise when not handled carefully, and scenario's to be considered. These will be discussed in the next chapters.

# Sending documents

## Discovery

In order for a sender to automatically use the Peppol network for the exchange of documents a discovery phase is necessary, to find out whether the recipient is present on the network, and to find out whether they can process the type of document wishes to send.

Based on data present in the document or the sender's ERP system, this information can be searched for on the Peppol network, simply by trying out each potential identifier of the recipient, and performing an SMP lookup until a suitable receiver endpoint is found.

Fully automatic discovery, where a specific bilateral agreement is not made regarding the identifier to use for sending documents, relies on the recipient to support the identifiers that are found in such a discovery process, see chapter 4 for discussion on this.

## Data points for discovery

Before constructing a document, the sender should always have some information about the recipient, usually stored in their ERP system. They usually have an existing business relationship, and there are legal requirements regarding the information that is contained in an invoice. Some of this information could be used to determine whether the invoice can be sent electronically over the Peppol network.

Within the Netherlands, the only legally required information about a recipient in an invoice is their legal name and address. Neither of these can be directly used to find the recipient on the Peppol network. But there are a number of likely data points that can be used. The most common of these

are:

* Legal identifier of the organization
* VAT number

Less often, but also used, are the following:

* IBAN number
* GLN number

In business to business relationships the legal identifier and VAT number are often known to the sender, and exchanged when forming a business relationship, or otherwise easily found. Other types are generally not known to the sender, and may not even be used by the recipient in the first place.

If known, all potential identifiers of the recipient could be searched in order and subsequently used as the recipient's address on the network. Note that this carries a heavy assumption, which shall be discussed in section 4.2.

## Data enrichment from external sources

One way to further help automatic discovery is to enrich known data from public sources, such as Chamber of Commerce registers, or the Peppol Directory.

This could be a dangerous proposition; multiple organizations with the same name might exist, or the name in the register may differ slightly from the name in a sender's ERP system.

Invoices that are sent to the wrong recipient by a system that performs automatic discovery result in lost invoices, and worse, leak potentially confidential information.

Extreme care should be taken when enriching the data with information from other sources. It might still be useful, but should at the least stop if there are ambiguous results, and ideally, include some form of additional verification of the data.

## Generation of Peppol BIS documents

In the case of Peppol BIS documents, the identifier that is the result of such a lookup must be present in the document itself (in the cbc:EndpointID element).

Therefore, this discovery phase must be done before the final XML is generated, and should either be performed by the software that creates the final XML, or provided as an extra service integration to that software.

Note the use of the word final; scenarios where the EndpointID is tentatively filled in by the software creating some intermediate form of the XML is possible, then finalized at a later step in the process are possible. But earlier discovery might be useful as well, if the software is capable of generating multiple formats, since this also provides insight in which document types are supported by the receiver.

If the discovery phase is performed by corner 2, it should only do so when corner 1 did not specify a specific identifier to send the document to. Corner 1 should always have an option to override automatic discovery, for scenarios where the recipient did communicate a specific identifier for receiving documents.

## When discovery is not possible

Sometimes, automatic discovery is simply not possible, even if both the sender and the recipient are present on the Peppol network. For instance, the recipient might use an identifier which is not known to the sender, or it might use a specific identifier (such as a GLN) for a particular department, and refuse documents for this department on the endpoint pointed to by the organization’s main legal registration identifier.

In such cases, some additional information still needs to be bilaterally exchanged and manually provided by the sender. It is therefore important that any system that uses automatic discovery has a fallback option where the sender can enter such information, either to override discovery, or as a fallback option when discovery fails.

# Receiving documents

## Publishing recipients

In order to be discovered automatically by (potential) senders, a recipient must be published in an SMP with an identifier that those senders know about. Since the goal of automatic discovery is to reduce the number of bilaterally agreed details, this means the identifier should be something that is either already known to the sender, or that needs to be communicated anyway.

It is therefore recommended to publish both the legal identifier and the VAT number of recipients in order to support automatic discovery at potential senders, as these are often exchanged when forming business relationships.

## Receiving and processing documents

If a receiver publishes support for a document type with different identifiers, all of which may be known to the sender, it is not possible to predict which of those identifiers might be used by senders that perform autodiscovery.

Therefore, in an autodiscovery scenario, it is vital that the handling on the receiving end does not depend on which identifier was actually used.

For instance, suppose a recipient is published by their VAT number, their legal identifier, they should not require some senders to use the VAT number, and others to use the legal identifier. Both should work for all senders.

## Routing based on recipient identifier

The previous section poses a potential problem for scenarios where internal routing at corner 3 in the 4-corner model is based on the recipient identifier. For instance, a large company may have their legal identifier published on the network, but they also have multiple branches, each of which has their own GLN, and their own service provider.

In such cases, processing invoices sent to a specific branch might fail in processing if they are sent to the legal identifier, and not the GLN of that branch.

It could be argued that a workaround for that problem would be to only publish recipients with GLN numbers, and not publish any other identifier types. However, this also poses several problems. For one, the use of specific identifier types is mandatory in some sectors or jurisdictions. Secondly, GLN numbers themselves could point to different levels of the organization, and multiple GLN numbers may be known to senders. Thirdly, this means automatic discovery is likely to fail, and bilateral exchange of additional data is again necessary for the correct functioning of the Peppol network.

This scenario is incompatible with automatic discovery, and the recipient should be very clear that it does not support it, and which identifier should be used to potential senders.

# Conclusion

Correct functioning of automatic discovery is an important next step for further automation and expansion of the Peppol network, which is beneficial for all parties involved.

We have discussed a number of potential issues regarding automatic discovery on the Peppol network, for both senders and receivers.

We recommend that these considerations are taken into account when designing and implementing systems that are connected to the Peppol network.