Belastingdienst

# IUC21-027 - Market Consultation

# Understanding Unstructured Files

Contracting authority(ies):

- Procurement Implementation Centre – Dutch Tax Administration (Inkoop Uitvoeringscentrum – Belastingdienst)

Version : 1.0
Reference : IUC21-027
Date : 9 July 2021

## Inhoud

# Chapter 1. General

## 1.1. Reading guide

Chapter 1 contains general information about the market consultation and information about the organisation of the Dutch Tax Administration (Belastingdienst) and the Procurement Implementation Centre (Inkoop Uitvoeringscentrum (IUC)).
Chapter 2 describes the desired situation and also contains questions.
Chapter 3, 'Market consultation procedure', sets out the procedure. It contains the information needed to participate in the market consultation.
Finally, a glossary can be found in Appendix 1.

## 1.2. Introduction

A market consultation is an organised exchange of information through questions and answers with interested parties about the intended contract. It is used by the Dutch Tax Administration to assess the feasibility and determine the preconditions. This market consultation is a preparation for the call for tenders Tooling Understanding Unstructured Files (Inzicht in Ongestructureerde Bestanden (IOB)).

The market consultation enables the Dutch Tax Administration to obtain knowledge and experience from the market and to test its own ideas for the assignment. It also improves the quality of the assignment documents.
Participation in the market consultation is entirely voluntary and without obligation. In addition, it has no consequences whatsoever for possible participation in the follow-up activities.

The Dutch Tax Administration has insufficient insight into the market's potential in the field of Solutios that can be used to gain understanding of the Dutch Tax Administration's unstructured files. It is desirable to use the market's knowledge to arrive at a qualitative market demand. To this end the Dutch Tax Administration invites market parties to respond. The contributions from the market are greatly appreciated. The market consultation ensures that a future assignment is well matched to the opportunities offered by the market.

The aim of the market consultation is to:

- raise interest among the parties in the market;
- determine the potential of the parties in the market (market forces, competition and cooperation between the market parties;
- determine the contract strategy (market approach, content and preconditions).

## 1.3. The Dutch Tax Administration

The Dutch Tax Administration is part of the Ministry of Finance which, in turn, is part of the legal entity the State of the Netherlands. The Dutch Tax Administration also includes:

- Customs (Douane): responsible for supervising the import, export and transit of goods through the Netherlands;
- Supplements (Toeslagen): responsible for the payment of income-related supplements;
- Fiscal Intelligence and Investigation Service (Fiscale Inlichtingen- en OpsporingsDienst (FIOD)).

The most well-known task of the Dutch Tax Administration is to levy and collect taxes and contributions. The Dutch Tax Administration not only collects, but also pays out. For example, the Dutch Tax Administration handles the payment of income-related supplements for children, childcare, rent and care.

Other important tasks are:
- detecting tax, economic and financial fraud;
- supervising the import, export and transit of goods;
- monitoring compliance with tax laws and regulations.

### 1.3.1. Procurement Implementation Centre - Dutch Tax Administration (IUC)

The IUC will supervise this market consultation for the contract. It is part of The Dutch Tax Administration SSO Facility Services (Belastingdienst / Shared Service Organisatie – Centrum voor Facilitaire Dienstverlening (B/SSO-CFD)) and carries out generic purchasing activities for the central government categories that have been placed under its management. It also handles specific procurement for the Dutch Tax Administration and Domains of Movable Property (Domeinen Roerende Zaken). For more information about the organisation of the Dutch Tax Administration, see the websites www.belastingdienst.nl and www.rijksoverheid.nl/ministeries/fin.

# Chapter 2. Desired situation

The Dutch Tax Administration considers it important to comply with statutory directives, such as: General Data Protection Regulation, Open Government Act and the Archives Act. The Dutch Tax Administration also sets an example in this respect.

In order to be able to manage these compliance issues even better, the Dutch Tax Administration intends to acquire a Solution that will provide insight into the Dutch Tax Administration's unstructured files. These insights can then be used to answer questions about unstructured files, thereby supporting the Dutch Tax Administration's business units, in particular, in managing compliance. Other important applications are gaining insight into archive-worthy documents for the Information Management department, and from the IT side it is desirable to gain more detailed insight into storage.

This Solution will be incorporated in a new ICT-service to be set up:
Understanding Unstructured Files (Inzicht in Ongestructureerde Bestanden (IOB)), which the Dutch Administration wishes to make available from its Government Data Centre (OverheidsDataCenter (ODC-BD)) to its clients, the Dutch Tax Administration (including the Directorates of Dutch Tax Administration, Supplements and Customs), as well as other authorities that make use of the services of ODC-BD.
The IOB service will then be able to provide both generic and specific overviews of the unstructured files, the number of which is currently estimated at about 1.5 billion files. The Lotus Notes mail environment comprises 28 servers with approximately 120 TB of data. For file storage, both an Isilon NAS application and 175 regular Windows file servers are used. The NAS application is mainly used for the storage of data accessible only to individual employees, which amounts to about 250 TB. The file servers are used for the storage of data from collaborative areas and this is about 150 TB. In addition, an IBM Connections environment of 20 TB is present.

The main aims of the IOB-service are:

- collecting information (metadata) about files on source systems;
- enriching this metadata on the basis of specific entity extraction, if necessary;
- central storage of this (enriched) metadata;
- being able to answer questions about the metadata in the form of dynamic dashboards and reports.

The Solution does not have to consist solely of software. It is possible that combinations of hardware, software and services are offered in the market. The Dutch Tax The Solution does not have to consist solely of software. It is possible that combinations of hardware, software and services are offered in the market. The Tax and Customs Administration wishes to receive information about all possible Solutions. Administration wishes to receive information about all possible Solutions.

In this market consultation, the following term (with a capital letter) and corresponding definition is used:
Solution: Subject of this market consultation. The set of systems (servers), software and services (e.g. support) that forms a working whole.

## 2.1. Requirements

The Solution is subject to the following requirements:

- The Solution is running and the data are located within the Tax Administration's data centres (on premise).
- The Solution must have Active Directory (AD) integration, for example LDAP that allows authentication and authorisation based on AD groups.
- The Solution must allow scanning of meta-data from unstructured data (files).

- The Solution must be able to scan characteristics (patterns) within file strings and store them as metadata.
- The Solution must be able to produce simple reports based on selections.

## 2.2. Wishes

With regard to the Solution the following wishes exist:

- The Solution must be based on selections simple - The Solution can be implemented on internally available hosting platforms (VMware, Windows or Redhat).
- The Solution can scan unstructured content in text fields (comments) of databases.
- The Solution has both a central scan functionality and the possibility of scanning specific data areas (selection of user(s)).
- The Solution must allow for advanced and customisable dashboards, Visual analytics and reporting (examples click-through dashboards, sphere diagram, pie charts).
- The Solution can track metadata over time, allowing for time-based reporting (time travel).
- The Solution has the ability to take actions (delete or move to other systems) based on selections.

## 2.3. Source systems

The following source systems / workloads are considered for the Solution:

- Files on file systems, specifically: NTFS, SMB (CIFS), Isilon NAS, local hard disk (C:) and in the future Object Based with file (SMB/NFS) access;
- Files/attachments in mail (IBM Notes);
- Files on collaboration platforms (IBM Connections);
- Office 365 files
- IBM Filenet

## 2.4. Questions - General

| | |
|---|---|
| Question 1 | Describe your Solution in up to 4 pages (format A4, Arial 9.5, line spacing single, margins 2.5). |

| | |
|---|---|
| Question 2 | Describe your support model in up to 2 pages (size A4, Arial 9.5, single line spacing, margins 2.5). |

| | |
|---|---|
| Question 3 | How is the Solution provided in the market (software only or vendor appliance)?<br>Substantiate the choice. What are the advantages and disadvantages of the Solution? |

| | |
|---|---|
| Question 4 | What specific features are present within the platform and why are they distinctive from the competition? |

| | |
|---|---|
| Question 5 | Which, preferably Dutch, references do you have that are comparable with the requested Solution? Please describe them and is it possible to contact these references? |

## 2.5. Data collection (scanning)

| | |
|---|---|
| Question 6 | How is metadata about the unstructured data collected? In what way are agents on the source systems used? |

| | |
|---|---|
| Question 7 | List the source systems from which metadata can be obtained? How do you deal with source systems from which metadata cannot be obtained at present? |

| | |
|---|---|
| Question 8 | List the types of files from which metadata can be obtained. Think of the intended source systems, but also binary files (e.g. photos) and compressed files. How do you deal with types of files for which metadata cannot be obtained at present? |

| | |
|---|---|
| Question 9 | How can your Solution search files based on patterns and enrich the metadata with the search results? |

| | |
|---|---|
| Question 10 | How can scanning activity be limited so that performance for regular use is maintained? |

| | |
|---|---|
| Question 11 | How is your Solution able to perform incremental scans? |

| | |
|---|---|
| Question 12 | Describe how and according to what rules scan data is kept and if necessary deleted? |

| | |
|---|---|
| Question 13 | Assuming 1.5 billion files and 1 petabyte of unstructured data, what is the approximate time it takes for a full and incremental scan (10% change rate) to collect metadata? |

| | |
|---|---|
| Question 14 | Does your product have a limit for the number of files and/or content items it can scan and if so, which ones and the possibilities to overcome these limits. |

| | |
|---|---|
| Question 15 | Can multi-tenancy be realised within your Solution? Think for example of splitting metadata collections of: Dutch Tax Administration, Supplements and Customs. |

| | |
|---|---|
| Question 16 | How does your Solution deal with encrypted files and what Solutions are offered? |

## 2.6. Datahygiëne/ -processing

| | |
|---|---|
| Question 17 | What possibilities are there to clean the scanned metadata from unwanted information or pollution? |

| | |
|---|---|
| Question 18 | What possibilities are there to enrich the metadata with extra attributes, for example hashing? |

| | |
|---|---|
| Question 19 | What possibilities are there to add business information. Think for example of organisation and personal data? |

| | |
|---|---|
| Question 20 | What possibilities are there for users or administrators to additionally enrich parts of the metadata with specific metadata? |

| | |
|---|---|
| Question 21 | To what extent does your Solution have self-learning capabilities, a self-learning thesaurus? |

| | |
|---|---|
| Question 22 | What possibilities are offered for finding files based on a defined set of characteristics (a pattern) that can be captured in a template and describe them? |

| | |
|---|---|
| Question 23 | Are standard templates, such as an AVG template, part of the Solution and if so, which templates are available and describe them.<br>Can the Tax and Customs Administration also create new templates itself within your Solution and what options and/or support are available? |

| | |
|---|---|
| Question 24 | What are the possibilities to follow user-initiated (clean-up) actions? For example, cleaning up a specific data area? |

| | |
|---|---|
| Question 25 | To what extent does your Solution have functionality to function as a link in a workflow? |

## 2.7. Visual analytics / reporting

| | |
|---|---|
| Question 26 | To what extent can reports about the collected metadata be designed and built by the user? Which functionalities are offered? Think for example of filtering and specific data segments. |

| | |
|---|---|
| Question 27 | To what extent can graphical presentations of the collected metadata be designed and built? Which functionalities are offered here? For example, the ability to click through on a data item for additional underlying information. |

| | |
|---|---|
| Question 28 | What graphical options does your Solution offer for presenting the metadata? |

| | |
|---|---|
| Question 29 | To what extent can the collected metadata be exported to a separate Visual Analytics Solution, for example SAS BI? |

## 2.8. Automation

| | |
|---|---|
| Question 30 | How can ad hoc or automated actions be performed on the basis of scan results? Think for example of the automatic digital archiving (in another Solution) of files of a certain age. |

| | |
|---|---|
| Question 31 | How can activities be planned? For example, the periodic refresh of scan data, running of reports. |

| | |
|---|---|
| Question 32 | Can mail be sent from your Solution, e.g. periodic reports? |

## 2.9. Beheer

| | |
|---|---|
| Question 33 | Can you describe in general terms what management functionality is included in your Solution? |

| | |
|---|---|
| Question 34 | What granularity in authorisation is possible in relation to the various functions? For example, can users use the system for their own job? Obviously only over the metadata of their personal files. |

| | |
|---|---|
| Question 35 | Can you describe the authorisation model of your Solution? Does it work with roles and tasks (RBAC) and is coupling with existing AD possible in this respect? |

| | |
|---|---|
| Question 36 | Describe the incident handling interface. Can the user interface display the name and location of the files, as well as who loaded the files and when?<br>Describe the capabilities of the software that include displaying a fragment of text in which the incidents occurred.<br><br>Describe the possibilities of creating summary reports, including number and type of incidents. What options are offered for filtering incidents by location or date? And what options are offered for exporting this data? |

| | |
|---|---|
| Question 37 | What opportunities do you offer to train users and administrators for your Solution? |

## 2.10. Hosting

| | |
|---|---|
| Question 38 | Can you describe in general terms how your Solution can be hosted and on the basis of which best practices this is done? |

| | |
|---|---|
| Question 39 | What are the system requirements of your Solution assuming metadata on 1.5 billion files and what will this system look like in 5 years with a growth of 25% per year? |

## 2.11. Auditing

| | |
|---|---|
| Question 40 | How does your system record certain actions and can these be monitored afterwards, whereby it must not be possible to mutate this record of actions? |

## 2.12. Security

| | |
|---|---|
| Question 41 | How does your Solution provide security functionalities, so that the personal metadata is maximally and demonstrably secure and not accessible to unauthorised persons? |

| | |
|---|---|
| Question 42 | What procedures and/or mechanisms does your organisation have in place to ensure that vulnerabilities related to this product are addressed in a timely manner and describe these? |

## 2.13. Licence

| | |
|---|---|
| Question 43 | What licensing models do you use and what are their advantages and disadvantages? |

| | |
|---|---|
| Question 44 | What is the cost structure of the licensing model? |

## 2.14. Other

| | |
|---|---|
| Question 45 | Describe the possibilities that your Solution offers that you think are important but which the Dutch Tax Administration has not asked for? |

# Chapter 4. Market Consultation Procedure

## 4.1. General

Any interested entrepreneur may participate in this market consultation. Participants in this market consultation are asked to answer the questions posed in chapter 3 and to provide an outline of the Solution to be used.

Written responses can be submitted by e-mail to hja.klein.ganseij@nelastingdienst.nl with reference to IUC21-027 IOB.
We would like to receive your contribution to the market consultation as soon as possible but no later than 11-8-2021, 24:00 hours, via TenderNed.

## 4.2. Contact

The Dutch Tax Administration has clear rules for this formal market consultation, therefore you should only communicate about this market consultation with the contact persons below, using the e-mail address below. Do you communicate about this market consultation with other employees of the or try to influence them via other channels. This may be a reason for the Dutch Tax Administration to exclude you from further participation in this market consultation..

| **Contact person** | |
|---|---|
| Name: | Ton (H.J.A.) Klein Ganseij |
| Function: | Senior Buyer ICT - IUC Dutch Tax Administration |
| E-mail: | hja.klein.ganseij@belastingdienst.nl |

## 4.3. Market consultation document

This document enables market parties to respond to the questions posed in this market consultation. The Dutch Tax Administration has described the need for this in this document and explained the procedure followed during the market consultation.

If, despite the care taken by the Dutch Tax Administration to prepare this document, you should come across any ambiguities or inadequacies or should have any questions, please let us know by sending an e-mail to hja.klein.ganseij@belastingdienst.nl before midnight on 23 July 2021, 24:00 hours. Anonymised answers will then be published in a memorandum of information.

The results of the market consultation may be included in tender documents and/or published, without being traceable to the participants in the market consultation. The Dutch Tax Administration may freely dispose of and use the information provided. We request that you answer the questions in writing. The information you provide will be treated confidentially. You may be approached to present your product at a (virtual) meeting with representatives of the Dutch Tax Administration.

## 4.4. Planning market consultation

| **Activiteit** | **Datum** |
|---|---|
| Publication market consultation on TenderNed | 9-7-2021 |
| Deadline for asking questions | 23-7-2021; 24:00 hours |
| Publication of any information note | 30-7-2021 |

| | |
|---|---|
| Deadline for sending in reply | 11-8-2021; 24:00 hours |
| Optional: oral explanations and supplementary questions | In consultation |

## 4.5. Oral explanation and supplementary questions

The Dutch Tax Administration may invite the participants in this market consultation to a digital meeting during which the invited participant can further explain its answers and during which the Dutch Tax Administration may ask additional questions in response to the responses received.

## 4.6. Form requirements

It is explicitly not the intention to submit tenders; we urge you to limit yourself to answering the questions posed in Section 3.
Please answer briefly and clearly. Naturally, you are not obliged to provide answers and you are also free to skip certain questions. The contribution to the market consultation/answering of questions should preferably be submitted in MS Word or Excel format, in the order of the questions. Additional and other documentation may be submitted as a pdf file. In the letter accompanying the contribution to the market consultation we would like to receive the following information from the contact person on the market parties' side: name, position, office address, telephone number and e-mail address.

| **Gegevens** | |
|---|---|
| Name | |
| Function | |
| Business adress | |
| Phone number | |
| Email adress | |

## 4.7. Preconditions for participation in this market consultation

By participating in this market consultation, you unconditionally agree to all the conditions as stated in this market consultation document. The conditions are listed exhaustively below:

- Participation in the market consultation is without obligation for market parties and the Dutch Tax Administration;
- This market consultation document is intended solely for market consultation purposes;
- The market consultation is not a part of the possible tendering procedure but a consultation round;
- The participants in this market consultation may not charge the Dutch Tax Administration any costs for the response to this market consultation and for the work to be carried out in this respect;
- This market consultation is explicitly not intended to make a preselection of market parties for a possible future tender procedure. Market parties that do not participate in the market consultation are therefore not excluded from participating in a possible tendering procedure. Neither are market parties who participate in the market consultation in any way excluded from, or favoured over, possible participation in a tendering or purchasing process;
- The results of the market consultation may be included in tender documents (e.g. by including information received during the market consultation in the Descriptive

Document), without it being possible to trace them back to the participants in the market consultation. By participating in the market consultation you agree with this;

- The information provided during the market consultation may deviate from the information to be provided during the tendering procedure;
- The input provided by participants in this market consultation shall be treated confidentially to the extent possible, taking into account in any case the legitimate interests of the participants in this market consultation and the Dutch Tax Administration;
- The language used during this market consultation can be either Dutch or English;
- The Dutch Tax Administration is in no way bound by the results of the market consultation;
- Participants in the market consultation cannot derive any rights from participating in the market consultation or from any information provided during the market consultation;
- It is not possible to accept or honour any conditions or claims (including conditions or claims relating to Intellectual Property rights such as copyright) of participants in this market consultation regarding the use of information provided by them to the Dutch Tax Administration within the context of this market consultation, or the confidentiality and secrecy of information provided by the market player in question. However, when providing, disclosing or processing information obtained from participants during, or as a result of, this market consultation, the Dutch Tax Administration shall, as far as possible, endeavour to prevent this information from being directly traceable to a particular participant in the market consultation.
- If intellectual property rights (including copyrights) of a participant or of third parties apply, the Dutch Tax Administration shall be indemnified by the participant in question against a claim from that participant and/or third parties on the basis of those intellectual property rights (including copyrights).
- The Dutch Tax Administration expressly reserves the right to do so: To change the planning of this market consultation as included in this document. In addition, to suspend the market consultation temporarily and/or permanently.

# Appendix 1 Glossary

| Afkorting | Uitleg |
|---|---|
| AD | Active Directory |
| AVG | Algemene Verordening Gegevensbescherming<br>General Data Protection Regulation |
| B/SSO-CFD | Belastingdienst SSO Facilitaire Dienstverlening<br>The Dutch Tax Administration SSO Facility Services |
| CIFS | Common Internet File System |
| FIOD | Fiscale inlichtingen- en opsporingsdienst.<br>Fiscal Intelligence and Investigation Service |
| IOB | Inzicht in ongestructureerde bestanden.<br>Understanding Unstructured Files |
| IUC | Inkoop Uitvoeringscentrum Belastingdienst<br>Procurement Implementation Centre – Dutch Tax Administration |
| LDAP | Lightweight Directory Access Protocol |
| NFS | Network File System |
| NTFS | New Technology File System |
| RBAC | Role Based Access Control |
| Solution | Subject of this market consultation. The set of systems (servers), software and services (e.g. support) that forms a working whole. |
| SAS-BI | Statistical Analysis System - Business Intelligence |
| SMB | Server Message Block |
| SSO | Shared Service Organisatie<br>Shared Service Organisation |
| Tooling | All necessary software for the IOB service. |
| WOO | Wet Open Overheid<br>Open Government Act |