Intelligente datamining voor auditdoeleinden – *iDM@audit*

Eindverslag

Programma	VLAIO - PIO
Aanbestedende diensten	AUDIT VLAANDEREN
Externe begeleider	ADDESTINO

Table of Contents

1 Context & Objectieven

1.1 Projectinitiators

1.1.1 Audit Vlaanderen

Audit Vlaanderen (AV) is een agentschap van de Vlaamse Overheid dat verantwoordelijk is voor het evalueren van de organisatiebeheersing (het internecontrolesysteem) van zowel de Vlaamse Administratie als de lokale besturen. Via het bieden van assurance, (beperkt) advies, kennisdeling en sensibilisering over risicomanagement, wil Audit Vlaanderen deze organisaties ondersteunen bij het beheersen van hun financiële, wettelijke en organisatorische risico's, [1] Het agentschap rapporteert over de uitgevoerde activiteiten en legt verantwoording af aan het Auditcomité van de Vlaamse Administratie en het Auditcomité van de lokale besturen[2].

Audit Vlaanderen voert verschillende soorten audits uit, waaronder reguliere en forensische audits. Reguliere audits worden gepland in functie van de risico's en focussen op het beoordelen van de beheersing van processen en systemen[3]. Forensische audits worden uitgevoerd om fraude en andere onregelmatigheden te onderzoeken, en kunnen zowel op eigen initiatief als op verzoek van derden worden uitgevoerd.

1.1.2 Programma Innovatieve Overheidsopdrachten (PIO)

Het Programma Innovatieve Overheidsopdrachten (PIO) helpt de Vlaamse Overheid en de publieke sector om innovatieve oplossingen te ontwikkelen, testen en valideren via overheidsopdrachten[4]. PIO richt zich op publieke noden waarvoor geen kant-en-klare oplossingen op de markt zijn. Door samen te werken met ondernemingen, kennis- en onderzoekscentra, streeft PIO naar het creëren van oplossingen die bijdragen aan een performantere overheid, competitievere ondernemingen en het aanpakken van maatschappelijke uitdagingen.

PIO biedt Vlaamse publieke organisaties de mogelijkheid om innovatieve oplossingen te laten ontwikkelen en/of valideren. Dit kan door middel van cofinanciering, waarbij PIO maximaal 50% van de kosten dekt, mits de aankopen voldoende impact en innovatie beogen. Daarnaast stelt PIO ook middelen ter beschikking voor het voorbereiden van de aankopen en biedt het advies en begeleiding gedurende het hele aankooptraject. Het programma staat open voor alle thema's en sectoren, en verwelkomt zowel kleine als grote innovatietrajecten.

1.2 Voortraject "PIO Audit Vlaanderen – iDM@audit"

1.2.1 Context

Bij het uitvoeren van de audits (zowel regulier als forensisch) staan de auditoren zelf in voor dataverzameling, -cleaning en -analyse. De huidige manier van werken is echter zeer arbeidsintensief en kampt met verschillende aandachtspunten, waaronder de omvang van de datatransfers tussen geauditeerde entiteiten en Audit Vlaanderen, de vele manuele handelingen die nodig zijn voor de verwerking, het gebrek aan tooling om ongestructureerde data te analyseren, en de versnippering van beschikbare informatie.

De afgelopen jaren deed Audit Vlaanderen ervaring op met enkele e-discoverysystemen. Naar aanleiding van het aflopen eind 2024 van een contract daarvoor, nam Audit Vlaanderen het initiatief om te evalueren of een één-op-één vervanging van de huidige tool, dan wel de vervanging door een meer intelligente en innovatieve oplossing de beste volgende stap is.

De overweging van het inzetten van innovatieve tool(s) is gedreven door de betrachting om het werk van de auditoren efficiënter en gemakkelijker te maken (bijvoorbeeld door indexeren en referentiëren te vereenvoudigen) en om de efficiëntie en kwaliteit van de audits te verhogen (bv. door een test te kunnen uitvoeren op een volledige populatie in nagenoeg dezelfde tijd waarin normaal enkel een steekproef kan worden getest). De beoogde efficiëntiewinsten slaan dus zowel op het bekomen van hetzelfde resultaat met minder inspanningen als op het realiseren van een beter resultaat met gelijkaardige inspanningen.

Dergelijke innovatieve tool(s) kan/kunnen dan gebruikt worden voor datatransfer, data-analyse, het efficiënt organiseren van, doorzoeken van en refereren naar informatie, alsook het creëren van een tijdslijn en het aanleveren van conclusies die als input kunnen dienen voor het auditverslag. .

In de mate dat voor sommige van deze tools de inzet van artificiële intelligentie (AI) nodig is, was het uitdrukkelijk niet de bedoeling van dit project om de randvoorwaarden (bv. DPIA, FRIA, beheerssystemen, richtlijnen, opleiding, technische en overige organisatorische maatregelen) voor het gebruik daarvan uit te klaren. Daaromtrent worden immers al significante inspanningen geleverd in het kader van het project van de Vlaamse overheid voor de uitrol van Copilot bij de Vlaamse administratie. De concretisering hiervan zal bij eventuele vervolgtrajecten sowieso verder moeten worden opgenomen.

1.2.2 Scope

In een audit zijn steeds verschillende personen en instanties betrokken (zie Sectie 2.2.1). In dit PIO-voortraject worden de medewerkers van Audit Vlaanderen geïdentificeerd als voornaamste stakeholders, met focus op innovatieve oplossingen die de efficiëntie binnen hun dagelijkse werkwijze verhogen en de pijnpunten zoals beschreven in Sectie 1.2.1 aanpakken.

Het auditproces zelf kan onderverdeeld worden in volgende high-level processtappen:

1. *Collect*: het verzamelen van informatie op een gecentraliseerde en betrouwbare opslaglocatie;
2. *Process*: het organiseren, structureren en opschonen van de verzamelde informatie;

3. *Explore*: het doorzoeken van de beschikbare informatie, het centraliseren van (tussentijdse) onderzoeksresultaten en het bijhouden van de nodige verwijzingen/referenties;
4. *Insight generation*: het detecteren van patronen, verwerven van inzichten, en het genereren en stapsgewijs verfijnen van een observatieverslag dat als input dient voor verdere rapportering;
5. *Report*: het compileren van een finaal auditverslag met de bevindingen en benodigde referenties, om gedeeld te worden met relevante personen en instanties (bv. het parket).

Er is ook een transversale processtap, 'Audit management', die zich uitstrekt over het volledige auditproces. Deze stap omvat het bijhouden, opvolgen en coördineren van de status en (tussentijdse) resultaten van de lopende audits binnen Audit Vlaanderen.

Deze processtappen worden uitgebreider besproken in Sectie 2.2.2**Error! Reference source not found.**. Ze dienen als referentiekader om de verschillende deelstappen binnen het auditproces gedetailleerd in kaart te brengen.

Binnen het huidige auditproces van Audit Vlaanderen zijn er verschillende aandachtspunten (zie verder Sectie 2.2.2). Een eerste aandachtspunt is de overdracht van grote hoeveelheden data bij de aanvang van een audit. Daarnaast vereist de verwerking van aangeleverde (on)gestructureerde gegevens nog veel manuele inspanning, omdat deze vaak niet meteen geschikt zijn voor analyse. In de daaropvolgende stappen botsen auditors op de beperkingen van gefragmenteerde tools en het ontbreken van geïntegreerde zoekmogelijkheden over verschillende databronnen en -formaten heen. Tot slot is ook de rapporteringsfase erg arbeidsintensief door het manueel opstellen van goed onderbouwde rapporten, waarbij meerdere revisies door verschillende betrokkenen nodig zijn.

1.2.3 Doelstelling van voortraject

Dit PIO-voortraject heeft als doel de noden scherp te stellen via interviews en workshops, en de technologische haalbaarheid te toetsen door middel van een marktconsultatie, ter voorbereiding van een innovatieve aanbesteding. De uitkomst vertaalt zich naar zowel voorliggend publiek verslag als een vertrouwelijk advies om de scope en aanpak van de innovatieve aanbesteding scherp te stellen.

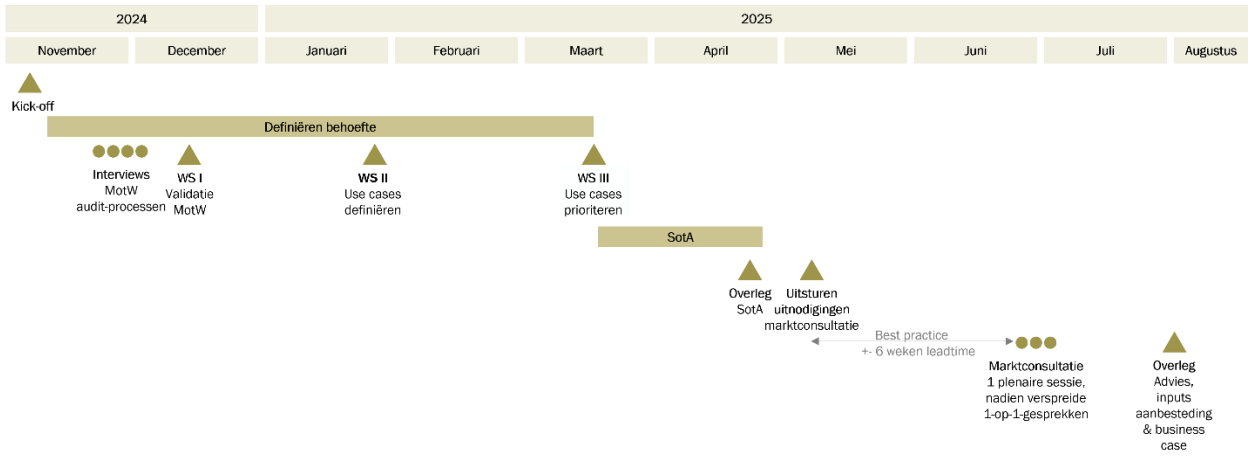
Het doel van dit PIO-voortraject is daarom ook tweeledig:

1. Zicht krijgen op de auditprocessen, om de bijhorende pijnpunten te begrijpen en om opportuniteiten te identificeren die kunnen leiden tot verbeteringen of verhoogde efficiëntie op vlak van databeheer- en verwerking;
2. Begrijpen wat de bereidheid van de markt is om te voorzien in de functionele en technische vereisten die volgen uit punt 1, gegeven de randvoorwaarden die bij deze auditfunctie in acht genomen moeten worden.

Dit voortraject heeft niet als doel om zich uit te spreken over de juridische aspecten die verbonden zijn aan het verzamelen en verwerken van gegevens binnen een auditcontext. Dit is evenwel een belangrijke randvoorwaarde (zie Sectie 2.2.4), die meegenomen dient te worden bij vervolgtrajecten.

1.3 Gevolgd proces en deelnemers

De tijdlijn voor de praktische planning van dit PIO-voortraject is weergegeven in Figuur 1.



Figuur 1: Plan van aanpak.

De belangrijkste stappen in het plan van aanpak worden hieronder uitgebreid beschreven:

1. Kick-off session
    - Doelstelling: definiëren van de scope van het project, het plan van aanpak en de eerste praktische aspecten (volgende stappen en tijdlijn)
    - Timing: donderdag 7 november, 2024
    - Locatie: Herman Teirlinckgebouw (Havenlaan 88, Brussel)
    - Deelnemers: 1 PIO-adviseur, 1 manager-auditor AV (forensisch), 1 expert-auditor AV, 2 auditoren AV (forensisch), 2 consultants van Addestino, 1 partner van Addestino
2. Interviews
    - Doelstelling: het in kaart brengen van de huidige auditprocessen, waarbij volgende vragen worden behandeld:
        - Welke databronnen worden er geraadpleegd?
        - Welken soorten data worden er geanalyseerd?
        - Over welke (data-)volumes gaat het?
        - Welke stappen en regels zijn er voor het behandelen van de auditdossiers?
        - Wie voert wat uit bij iedere stap in het proces?
        - Welke tools worden in elke stap gebruikt?
        - Wat zijn de huidige pijnpunten of niet-ingevulde noden?
    - Timing: dinsdag 19 november
    - Locatie: Herman Teirlinckgebouw (Havenlaan 88, Brussel)
    - Deelnemers: 7 auditors AV (4 forensisch, 2 regulier – Vlaamse Administratie, 1 regulier – lokale besturen), 1 expert-auditor AV, 2 consultants van Addestino
    - Output: op basis van deze interviews en bestaande documentatie werd een eerste versie van de huidige manier van werken gedocumenteerd in een “Map of the World” (MotW). Deze output dient als leidraad in de volgende stappen van het voortraject.

3. Workshop 1 (WS I) – Validatie MotW
    - Doelstelling: doorlichting van de eerste versie van de MotW en validatie door de betrokkenen van Audit Vlaanderen
    - Timing: maandag 16 december 2024 (13:30–15:00)
    - Locatie: Herman Teirlinckgebouw (Havenlaan 88, Brussel)
    - Deelnemers: 1 manager-auditor AV (forensisch), 1 expert-auditor AV, 4 auditoren AV (3 forensisch, 1 regulier – lokale besturen), 2 consultants van Addestino
    - Output: op basis van de feedback tijdens deze sessie werd een tweede iteratie van de MotW gecreëerd waarin de benodigde aanpassingen werden aangebracht.
    - Communicatie: het eerste ontwerp van MotW werd kort toegelicht op het managementteam (MAT) van 18 december 2024
4. Workshop 2 (WS II) – Opstellen *use cases*
    - Doelstelling: scherpstellen van de behoeften van de stakeholders (medewerkers Audit Vlaanderen) via het opstellen van *use cases*
    - Timing: vrijdag 31 januari 2025 (9:15–13:15)
    - Locatie: Herman Teirlinckgebouw (Havenlaan 88, Brussel)
    - Deelnemers: 1 manager-auditor AV (forensisch), 1 expert-auditor AV, 8 auditoren AV (5 forensisch, 3 regulier – Vlaamse Administratie, 1 regulier – lokale besturen), 1 PIO-adviseur, 2 consultants van Addestino
    - Output: een lijst met (nog niet geprioriteerde) *use cases*.
5. Workshop 3 (WS III) – Prioritering van de *use cases*
    - Doelstelling: valideren en gezamenlijk prioriteren van de *use cases* opgesteld in WS II
    - Timing: woensdag 19 maart 2025 (9:15–13:15)
    - Locatie: Herman Teirlinckgebouw (Havenlaan 88, Brussel)
    - Deelnemers: 1 manager-auditor AV (forensisch), 1 expert-auditor AV, 6 auditoren AV (4 forensisch, 1 regulier – Vlaamse Administratie, 1 regulier – lokale besturen), 1 PIO-adviseur, 2 consultants van Addestino
    - Output: een lijst met geprioriteerde *use cases*
6. Read-out van state-of-the-art analyse & referentiearchitectuur
    - Doelstelling: op basis van de geprioriteerde *use cases* en een state-of-the-art analyse stelt Addestino een referentiearchitectuur voor, die van belang zal zijn voor het bepalen van de relevante spelers voor de marktconsultatie. Typisch wordt de referentiearchitectuur opgesteld door te vertrekken vanuit drie verschillende invalshoeken.
    - Timing: donderdag 24 april 2025 (15:00–17:00)
    - Locatie: Online (Microsoft Teams)
    - Deelnemers: 1 expert-auditor AV, 2 consultants van Addestino
    - Output: verduidelijking van het voorstel aan het MAT tot opstart van een marktconsultatie
    - Communicatie: op 6 mei 2025 werd de stand van zaken van het PIO-project met het MAT besproken. Daartoe werden o.a. de MotW, de use cases, de geprioriteerde use cases, de situering van de use cases in het proces, een door Addestino opgesteld tussentijds rapport, voorstellen van Addestino omtrent de referentiearchitectuur en

bevindingen uit een desktop-marktanalyse door Addestino aan het MAT voorgelegd. Uit de bespreking bleek voldoende gedragenheid om tot de voorgestelde markconsultatie over te gaan.

7. Marktconsultatie met relevante spelers
    - Doelstelling: begrijpen wat de bereidheid is van de marktspelers om te voorzien in de vooropgestelde technische en functionele vereisten van de referentiearchitectuur
    - Timing: plenaire sessie op dinsdag 24 juni 2025 (14:00–17:00), één-op-één-gesprekken op donderdag 26 juni 2025 (14:00–16:00), vrijdag 17 juni 2025 (9:30–11:30) en vrijdag 4 juli 2025 (13:30–15:30)
    - Locatie: Online (Microsoft Teams)
    - Deelnemers: 1 expert-auditor AV, 1 auditor AV (forensisch), 1 PIO-adviseur, 2 consultants van Addestino en 34 bedrijven (zie lijst in Annex B)
8. Eindverslag
    - Doelstelling: Voorstellen van het eindverslag, met daarin een aanbeveling voor de volgende stappen op vlak van innovatieve aanbestedingen ter ondersteuning en verbetering van de auditprocessen.
    - Timing: augustus 2025

2 Analyse van de noden en randvoorwaarden

2.1 Gevolgd proces

Zoals beschreven in Sectie 0 werden voor het analyseren van de noden eerst enkele interviews georganiseerd met verschillende stakeholders binnen Audit Vlaanderen. In analogie met de scope van het PIO-voortraject lag de focus van deze gesprekken voor een groot deel op de forensische audits, maar er werd ook dieper ingegaan op aspecten die voor reguliere audits ook een meerwaarde zouden kunnen betekenen. Op basis van bestaande documentatie en een reeks van interviews werd de huidige manier van werken zo nauwkeurig mogelijk gedocumenteerd door Addestino, resulterend in een eerste versie van de MotW. In deze voorstelling wordt het auditproces gestructureerd volgens de high-level processtappen die beschreven werden in Sectie 1.2.2.

De initiële versie van de MotW werd door Addestino voorgesteld tijdens de eerste workshop (WS I), die plaatsvond op maandag 16 december 2024 in Brussel. Hier kregen de betrokkenen van Audit Vlaanderen de kans om de MotW te valideren, alsook om hun feedback, toevoegingen of bemerkingen te delen. Aan de hand van deze feedback werd vervolgens een finale versie van de MotW opgesteld door Addestino, die met alle stakeholders gedeeld werd. Deze MotW werd doorheen het verdere traject gebruikt als leidraad om de gebruikersbehoeften gedetailleerd in kaart te brengen.

De resultaten van deze analyse worden hieronder in detail beschreven.

2.2 Resultaten

2.2.1 Rollen

In het auditproces zijn verschillende actoren betrokken, die onder te verdelen zijn in drie categorieën. De eerste categorie omvat de personen verbonden aan de geauditeerde entiteit. In het geval van een forensische audit zijn binnen deze categorie twee groepen specifiek het vermelden waard. Dit zijn enerzijds de personen die het onderwerp vormen van de forensische audit zelf (de zogenaamde ‘custodians’), en anderzijds IT-experts die samen met Audit Vlaanderen zorgen voor de verzameling en overdracht van informatie. Onder de tweede categorie vallen de interne medewerkers van Audit Vlaanderen: (senior) auditoren, manager-auditoren en de administrateur-generaal. De auditoren werken gewoonlijk in teamverband aan een case. Manager-auditoren sturen de auditoren aan en finaliseren het auditverslag. De administrateur-generaal heeft de leiding over Audit Vlaanderen. De laatste categorie bevat de actoren die instaan voor externe ondersteuning doorheen het auditproces. Dit kan gaan over IT-ondersteuning of auditmatige ondersteuning (bijvoorbeeld externe auditoren die ingehuurd worden via een raamcontract), maar ook overheidsdiensten die als externe bron informatie aanleveren (bijvoorbeeld de Kruispuntbank van Ondernemingen of het Omgevingsloket) maken hier deel van uit. Een overzicht van deze verschillende rollen, onderverdeeld in de hierboven beschreven categorieën, is weergegeven in Figuur 2.

Figuur 2: Betrokkenen in het auditproces, onderverdeeld in drie categorieën.

2.2.2 Auditproces

De high-level processtappen die gedefinieerd werden in Sectie 1.2.2 vormen de kapstok voor de structuur van het volledige auditproces, zoals weergegeven in de MotW (Figuur 3). Hieronder worden alle aspecten binnen elk van deze processtappen beschreven (hier wordt uitgegaan van een forensische audit, in het geval van een reguliere audit zijn een aantal stappen niet van toepassing en worden sommige stappen technisch anders uitgevoerd):

Collect

De collectiefase start bij forensische audits altijd met een melding (bv. van een klokkenluider). Na een initiële beoordeling, waarbij wordt nagegaan of Audit Vlaanderen bevoegd is om de audit uit te voeren en of er voldoende aanwijzingen zijn van intentionele en mogelijk straf- of tuchtrechtelijke inbreuken, wordt gestart met het verzamelen van informatie. Deze informatie omvat diverse bestandstypes, waaronder vooral documenten en e-mails, maar eventueel ook andere dossierstukken en/of communicatiegegevens. Gewone bestanden worden opgeladen in een gedeelde SharePoint-omgeving, hetzij rechtstreeks door de geauditeerde, hetzij via de auditoren die de bestanden ontvangen via deelruimtes, e-mail of een externe drive. Vervolgens worden de bestanden door de auditoren overgezet naar een interne SharePoint-omgeving. Voor deze stappen gaan auditors van Audit Vlaanderen doorgaans on-site bij de geauditeerde entiteit om de IT-verantwoordelijke te begeleiden bij het correct aanleveren van de gevraagde gegevens. Daarnaast verzamelen auditoren ook zelf bijkomende informatie uit externe bronnen, zoals de Kruispuntbank van Ondernemingen (KBO) of het Omgevingsloket. Deze informatie wordt manueel toegevoegd aan de interne bestandsomgeving.

Process

In deze fase wordt de verzamelde informatie klaargezet zodat er verder onderzoek op kan gebeuren. Bestanden worden gecategoriseerd en voorzien van labels om de verdere verwerking te vergemakkelijken. E-mailconversaties worden toegewezen aan de juiste onderzochte personen binnen de geauditeerde entiteit (zie Figuur 2) en vervolgens onderworpen aan een deduplicatieproces, zodat redundanties (in de e-mails zelf maar ook in bijlagen) vermeden worden.

Explore

De exploratiefase richt zich op het doorzoeken en analyseren van de verzamelde informatie. In bepaalde gevallen kunnen er zaken worden uitbesteed aan externe partijen voor gespecialiseerde analyses. Voor documenten gebeurt dit onderzoek meestal handmatig binnen de interne SharePoint-omgeving. Conversaties (bv. E-mails) en bijlagen worden doorzocht binnen de *e-discovery*-module 'Relativity' van PWC, waarbij indien nodig OCR-technologie wordt ingezet om tekst uit gescande documenten en afbeeldingen te extraheren. Relevante documenten en e-mails worden gelabeld (bv. als 'relevant'), waarna een lijst van deze gelabelde items wordt geëxporteerd. Tot slot worden de referenties naar de relevante documenten en conversaties samengebracht in een centraal Excel-bestand, dat als basis dient voor de volgende fase (*Insight generation*).

Insight generation

Op basis van de gerefereerde bewijsstukken in het centrale Excel-bestand uit de vorige processtap (*Explore*) worden inzichten gegenereerd, verbanden gelegd en patronen herkend. Hierbij worden visuele hulpmiddelen ingezet, zoals patronen en tijdslijnen. Interviews met *custodians* maken eveneens deel uit van deze fase en kunnen leiden tot nieuwe inzichten, waardoor eventueel opnieuw informatie moet worden verzameld en geanalyseerd. Dit iteratieve karakter is vaak kenmerkend voor de auditcyclus, die niet altijd als een puur lineair proces kan worden voorgesteld. Wanneer het auditdossier voldoende onderbouwd is, worden confrontatiegesprekken gevoerd, waarin onderzochte personen binnen de geauditeerde entiteit worden geconfronteerd met de bevindingen, en vervolgens de kans krijgen om weerwoord te bieden. De resultaten van het gevoerde onderzoek en de woordelijke notulering van deze gesprekken worden vastgelegd in een observatiedocument, dat als input dient voor de rapporteringsfase.

Report

De rapporteringsfase omvat het opstellen van een finaal verslag waarin alle bevindingen systematisch worden vastgelegd, en waarbij steeds correct wordt gerefereerd naar de relevante bewijsstukken. Dit gebeurt in meerdere stappen: eerst wordt een aanzet geformuleerd op basis van het observatiedocument uit de vorige processtap (*Insight generation*). In een eerste revisie door een senior auditor worden de hyperlinks naar de brondocumenten gecontroleerd. Vervolgens wordt bij een tweede revisie door een manager-auditor de tekst geherformuleerd, en worden de relevante (paragrafen/secties uit) bewijsstukken verbatim opgenomen. Tot slot ondergaat het rapport een laatste revisieronde door de administrateur-generaal, alvorens het wordt verspreid naar de bevoegde instanties.

Bij het analyseren van het huidige auditproces springen enkele zaken in het oog. Zo is er een duidelijke splitsing tussen het documentenonderzoek enerzijds en het mailboxonderzoek (soms ruimer: onderzoek van relevante conversaties) anderzijds. Voor elk van deze flows bestaat er een goed uitgewerkte methodiek, maar worden er andere tools gebruikt om de specifieke onderzoeksdoelstellingen te ondersteunen. Daarnaast wordt het auditproces ook gekenmerkt door een sterk iteratief karakter. Tijdens de uitvoering van de audit is er namelijk sprake van voortdurende terugkoppeling, bijsturing en verfijning van eerdere analyses, op basis van nieuwe bevindingen of aanvullende informatie (bijvoorbeeld uit de interviews).

Bij aanvang van dit voortraject stonden al enkele aandachtspunten op de radar bij de teams binnen Audit Vlaanderen. Deze aandachtspunten situeren zich in verschillende processtappen en worden hieronder kort beschreven.

Een eerste punt omvat het transfereren van data in de collectiestap. Om een audit uit te kunnen voeren, baseert Audit Vlaanderen zich namelijk vooral op informatie die verstrekt wordt door de geauditeerde entiteit zelf. Het kan daarbij gaan om gestructureerde data (bijvoorbeeld boekhoudkundige gegevens, overzichten van leveranciers, …), maar ook om niet-gestructureerde informatie (bijvoorbeeld overeenkomsten, verslagen, bestekken, offertes, …). Meestal gaat het om aanzienlijke datavolumes, die snel getransfereerd moeten kunnen worden naar een beveiligde opslagomgeving van (of beheerd in opdracht van) Audit Vlaanderen. De beschikbaarheid van een laagdrempelige, intuïtieve, performante en veilige manier om informatie te transfereren, is bepalend voor de snelheid en gebruiksvriendelijkheid van deze eerste stap in de audit. Uiteraard moet hierbij steeds ook voldoende aandacht besteed worden aan het garanderen van de integriteit (alsook veiligheid en vertrouwelijkheid) van de opslaglocatie, bijvoorbeeld door preventieve veiligheidsmaatregelen in te bouwen (zoals virusscans of een controle op corrupte data).

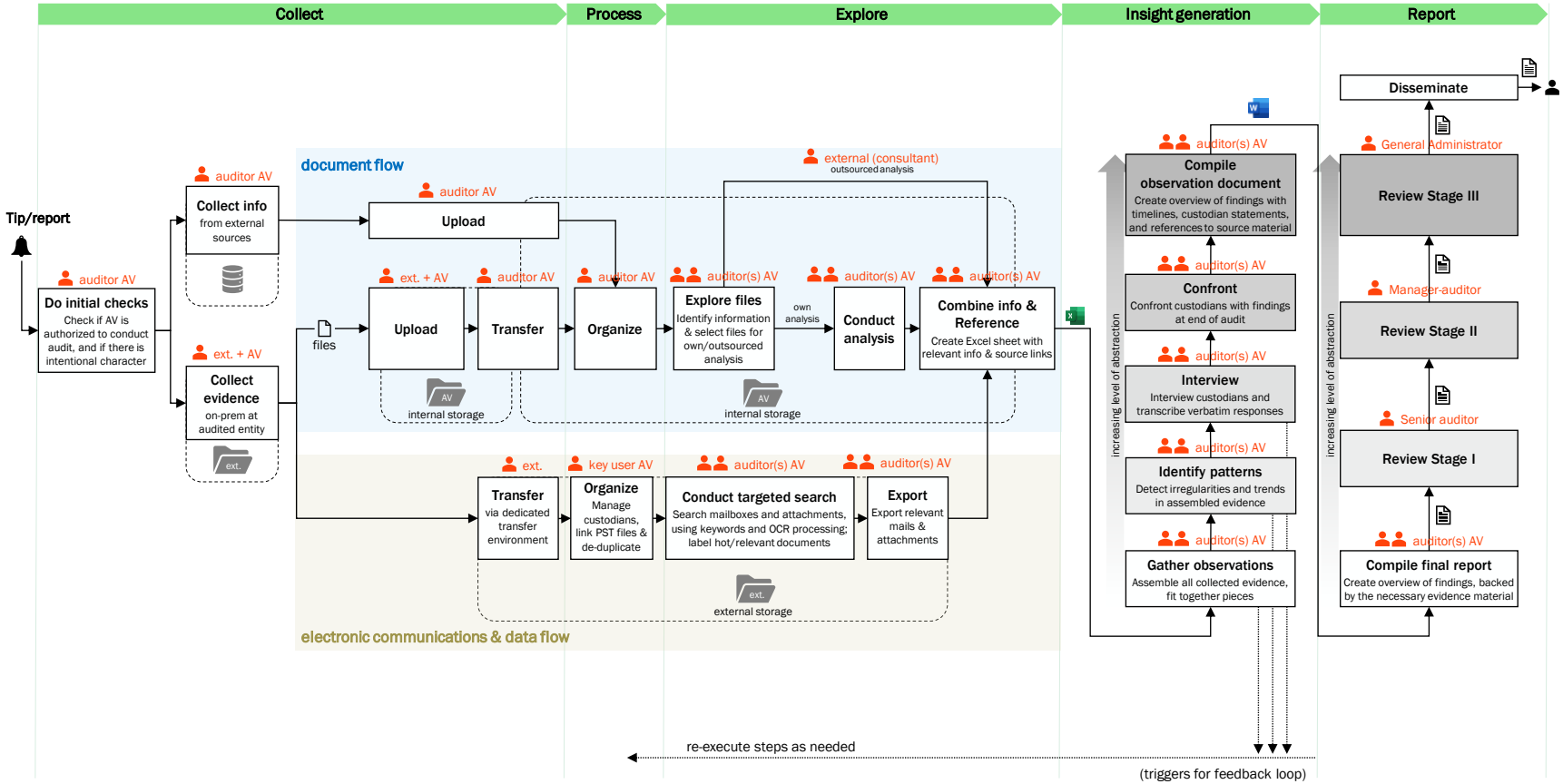
Een ander aandachtspunt situeert zich op niveau van de processing-stap. Veelal worden gestructureerde gegevens zoals boekhoudrapporten of inventarissen aangeleverd in een formaat dat niet onmiddellijk geschikt is voor data-analyse in functie van de onderzoeksvragen van de audit. Hierdoor moet de auditor de aangeleverde databestanden eerst herstructureren, samenvoegen en aanvullen, vooraleer ze bruikbaar zijn voor het doorzoeken van de informatie. Verdere automatisering kan ervoor zorgen dat deze stappen minder arbeidsintensief worden en kan er mogelijks voor zorgen dat fouten sneller gespot worden, waardoor de auditoren sneller kunnen opleveren en aan minstens dezelfde kwaliteit.

In de stappen rond *Explore* en *Insight generation* zitten de grootste uitdagingen op inhoudelijk vlak. Tijdens de data-analyse wordt informatie gezocht in verschillende informatiebronnen met uiteenlopende dataformaten. Momenteel gebeurt dit met verschillende tools, waardoor één integrale zoekactie, over alle databronnen heen, niet steeds mogelijk is. Bovendien zitten de resultaten van de verschillende analyses verspreid over verschillende tools en documenten. Het aanvullen van de tools voor analyses op ongestructureerde bronnen en het consolideren van de tools en informatie kan de efficiëntie van het auditproces verder verhogen.

Een laatste aandachtspunt situeert zich in de rapporteringsfase. Het opstellen van een consistent en goed onderbouwd rapport, waarbij correct wordt verwezen naar het juiste bewijsmateriaal, is momenteel een arbeidsintensief proces dat meerdere controle- en revisierondes vereist door verschillende betrokkenen. Het versnellen of stroomlijnen van deze procedure, bijvoorbeeld via de inzet van generatieve AI-technologieën die een eerste conceptversie van verwoordingen van de vaststellingen uit het observatiedocument kunnen aanleveren, kan aanzienlijke efficiëntiewinsten opleveren.

Hoewel de focus van dit PIO-voortraject op de forensische audits ligt, kan het nuttig zijn om de gelijkenissen en verschillen met reguliere audits goed voor ogen te houden. Sommige oplossingen die voorgesteld worden voor forensische audits kunnen immers evengoed relevant blijken voor de reguliere audits. Reguliere audits verlopen min of meer volgens dezelfde high-level processtappen zoals hierboven beschreven, maar er zijn ook enkele verschillen met forensische audits:

- Reguliere audits starten niet met een melding, maar worden uitgevoerd volgens een vooropgestelde planning. Deze planning komt tot stand op basis van een overkoepelende risicoanalyse op verschillende domeinen.
- Bij reguliere audits is er bij aanvang van de audit al een duidelijk overzicht van de scope, waar dit bij forensische audits niet het geval is. De op te vragen informatie en benodigde analyses zijn echter bij beide audittypes vaak afhankelijk van het auditverloop.
- Reguliere audits richten zich niet op mailboxonderzoek en maken dan ook geen gebruik van de tool 'Relativity'. Verder worden zowel bij reguliere als forensische audits documenten onderzocht en interviews/confrontatiegesprekken met verdachten uitgevoerd.
- Rapportering is bij alle audits afgestemd op het doelpubliek: dit betekent dat forensische audits op een andere manier rapporteren dan reguliere audits, en dat er bij die laatste categorie ook een verschil is tussen rapportage voor de Vlaamse Administratie en voor de lokale besturen.
- De doorlooptijd bij reguliere audits is doorgaans korter dan bij forensische audits.



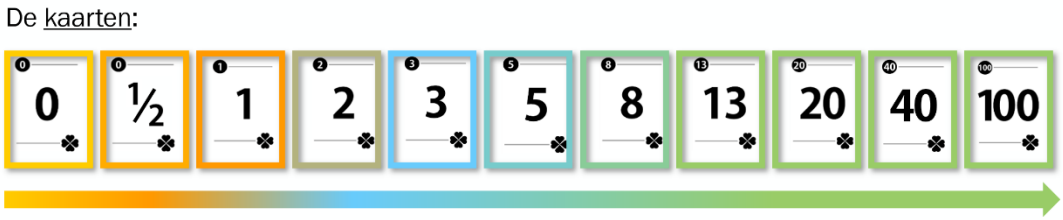
Figuur 3: Map of the World (MotW), die het (forensische) auditproces in kaart brengt.

2.2.3 Noden

Om de leesbaarheid te vergroten werden verschillende *use cases* gegroepeerd volgens de processtappen (zie Secties 1.2.2 en 2.2.2):

- *Use cases* met **prefix C** komen overeen met de '**Collect**' stap: het verzamelen van documenten en informatie van de geauditeerde entiteit.
- *Use cases* met **prefix P** komen overeen met de '**Process**' stap: het verwerken en organiseren/structureren van de verzamelde informatie.
- *Use cases* met **prefix E** komen overeen met de '**Explore**' stap: het doorzoeken van de beschikbare informatie.
- *Use cases* met **prefix I** komen overeen met de '**Insight generation**' stap: het genereren van inzichten op basis van het gevoerde onderzoek.
- *Use cases* met **prefix R** komen overeen met de '**Report**' stap: het opleveren van de conclusies in een finaal verslag.
- *Use cases* met **prefix M** komen overeen met de transversale '**Audit Management**' stap: het opvolgen en centraliseren van informatie rond alle lopende audits.

De toegevoegde waarde van elke *use case* wordt toegekend door middel van een *planning poker* techniek, waarbij elke *use case* een score krijgt die de toegevoegde waarde van de *use case* aangeeft. Onderstaande figuur geeft weer hoe de verschillende scores kunnen geïnterpreteerd worden:



Deze waardes werden tijdens workshops door de aanwezigen toegekend en gemotiveerd. Hieronder zijn een aantal *use cases* samengevoegd.

Use Case C.1: "Als auditor kan ik aan de hand van een inventaris/overzicht het verzamelen van documenten en informatie in gang zetten en opvolgen, zodat het collecteren van bewijsstukken efficiënt en zonder fouten verloopt"

Score van de toegevoegde waarde: **13**

Bij de opvolging van de informatieverzameling kan een zekere tijdswinst worden behaald, aangezien het contacteren van de geauditeerde entiteit voor ontbrekende informatie vaak een tijdrovend proces is. Ook op het gebied van *quality assurance* biedt dit een meerwaarde, doordat de kans kleiner wordt

dat te verzamelen informatie uit het oog wordt verloren. Het is echter belangrijk op te merken dat veel auditoren al een eigen opvolgingssysteem hanteren (bijvoorbeeld in Excel of via Outlook Calendar), waardoor de potentiële efficiëntiewinst beperkt is. De meerwaarde ligt dus vooral in het feit dat de opvolging veel gestroomlijnder kan verlopen met een uniforme tool voor de gehele organisatie.

***Use Case C.2**: "Als auditor kan ik reminders instellen voor de geauditeerde entiteit omtrent opgevraagde bestanden, zodat er geen manueel werk nodig is om de geauditeerde entiteit te contacteren bij laattijdige bezorging van opgevraagde bestanden"*

Score van de toegevoegde waarde: **besproken maar niet gescoord (meegenomen in C.1)**

***Use Case C.3**: "Als auditor kan ik suggesties krijgen van op te vragen (standaard)documenten voor een bepaalde auditcategorie (die ik zelf aangeef), zodat vaakvoorkomende documenttypes (volgens de auditcategorie) zeker niet over het hoofd gezien worden bij het opstellen van de inventaris van op te vragen documenten"*

Score van de toegevoegde waarde: **5**

De meerwaarde van deze *use case* wordt eerder laag ingeschat, omdat de auditoren vaak zeer goed weten welke documentatie al dan niet dient opgevraagd te worden in het kader van een specifieke audit.

***Use Case C.4**: "Als geauditeerde kan ik zelf metadata (inhoud, personen, data, samenvattende beschrijving, …) toevoegen aan opgeladen bestanden, zodat zij die de bestanden goed kennen (bij de bron) zoveel mogelijk context kunnen meegeven (ofwel: om te vermijden dat de auditors zelf meer tijd moeten steken in het opbouwen van deze context)"*

Score van de toegevoegde waarde: **8 (regulier) – 3 (forensisch)**

Zowel voor reguliere als forensische audits wordt de meerwaarde van deze *use case* als laag ingeschat. Er zou in principe beperkte tijdswinst kunnen worden gerealiseerd door de verantwoordelijkheid voor het aanleveren van goed gedocumenteerde data bij de geauditeerde zelf te leggen, maar dit veronderstelt een volledige en correcte medewerking van de geauditeerde. De geauditeerde entiteit kan, bewust of onbewust, incorrecte informatie als metadata aanleveren, waardoor de beoogde efficiëntiewinst verloren gaat. Bovendien is het voor forensische audits ongewenst om de geauditeerde entiteit inzicht te geven in welke metadata verwacht wordt, omdat dit mogelijke denkpistes kan onthullen. Voor reguliere audits is dit laatste argument minder relevant, waardoor de *use case* voor hen iets hoger gescoord wordt.

***Use Case C.5**: "Als auditor kan ik technische metadata van documenten/bestanden/e-mails behouden tijdens alle stappen van de verdere verwerking, zodat zo veel mogelijk context (hier: technische metadata) bij de bewijsstukken bewaard blijft en gebruikt kan worden om inzichten te verwerven en conclusies te staven"*

Score van de toegevoegde waarde: **besproken maar niet gescoord**

***Use Case C.6**: "Als auditor kan ik bij ontvangst van PDF's/foto's/etc. automatische bestandsnamen toekennen op basis van een scan op de inhoud, zodat onduidelijke bestandsnamen het auditproces niet vertragen"*

Score van de toegevoegde waarde: **13**

Deze *use case* wordt door de meerderheid van de auditoren niet als een absolute must gezien. De meerwaarde ervan hangt sterk af van de geanalyseerde dataset binnen de audit in kwestie: de benamingen van files in het bestandssysteem van de geauditeerde zelf zullen naar hoge waarschijnlijkheid intuïtief gekozen zijn, in tegenstelling tot bijvoorbeeld extracten uit boekhoudkundige systemen, waar meer cryptische benamingen verwacht worden.

***Use Case C.8**: "Als auditor kan ik juridische adviezen uit vorige audits doorzoeken (op basis van trefwoord, bron of met een AI-antwoord), zodat adviezen niet dubbel opgevraagd moeten worden"*

Score van de toegevoegde waarde: **20**

De mogelijkheid om in eerdere cases te zoeken naar juridische adviezen die relevant kunnen zijn voor de huidige case, kan zeker nuttig zijn en tijdswinst opleveren. Het ontbreken van deze functionaliteit zou echter geen dealbreaker zijn, aangezien het rondvragen bij collega's een laagdrempelige en werkbare oplossing biedt.

***Use Case C.9**: "Als auditor kan ik de geauditeerde entiteit contacteren via een gecentraliseerd communicatieplatform (bv. een toegewijd communicatiekanaal, of een interface die verschillende bestaande kanalen zoals telefoon/e-mail/fileshares centraliseert), zodat er geen communicatiekanalen uit het oog verloren worden tijdens het verzamelen van bestanden en informatie"*

Score van de toegevoegde waarde: **niet besproken**

***Use Case C.10**: "Als auditor kan ik bewijsstukken rechtstreeks ontvangen in een opslagomgeving waarin alle verdere verwerking gebeurt, zodat tussen het collecteren en (verschillende stappen van) verwerken geen tijd verloren gaat aan het kopiëren van grote volumes aan data."*

Score van de toegevoegde waarde: **niet besproken**

***Use Case C.11**: "Als deel van de geauditeerde entiteit kan ik enkel de eigen opgeladen bestanden zien, zodat bestaande scheidingen tussen departementen en rollen bij de geauditeerde partij behouden blijven op vlak van gegevensbescherming"*

Score van de toegevoegde waarde: **niet besproken**

***Use Case C.12**: "Als auditor kan ik toegang krijgen tot IT-systemen van de geauditeerde entiteit om zelf een forensische kopie van relevante data te maken, zodat er geen afhankelijkheid is van de IT-verantwoordelijke bij de geauditeerde entiteit (bv. wanneer deze weinig ervaren is waardoor het langer kan duren of niet van de eerste keer juist/volledig is)"*

Score van de toegevoegde waarde: **niet besproken**

***Use Case C.13**: "Als auditor kan ik door de geauditeerde externe drives of opslaglocaties laten opladen in de transferomgeving, zodat ook externe drives of opslaglocaties als een bron van informatie kunnen dienen en digitaal doorzocht kunnen worden"*

Score van de toegevoegde waarde: **niet besproken**

Use Case P.1: "Als auditor kan ik video's laten analyseren op inhoudelijke elementen (objecten, scenes en activiteiten) op basis van beeld en geluid, waarbij een oplijsting wordt gegenereerd van op welk tijdstip/frame zich iets voordoet, zodat relevante video's of stukken uit video's snel geïdentificeerd kunnen worden"

Score van de toegevoegde waarde: 5

In het kader van audits waarbij een grote hoeveelheid beeldmateriaal (bijvoorbeeld van bewakingscamera's) moet worden onderzocht, kan dit een aanzienlijke tijdswinst opleveren. Dergelijk onderzoek komt echter zelden voor, wat de globale tijdswinst in perspectief plaatst. Bovendien is de aard van de analyse sterk afhankelijk van de specifieke audit: dit kan variëren van gezichtsherkenning en objecttracking tot het detecteren van gedragingen of andere visuele patronen. Aangezien hier geen uniforme lijn in te trekken valt, is het onduidelijk in hoeverre een generieke functionaliteit effectief ondersteuning kan bieden. Om deze redenen werd de meerwaarde van deze *use case* als laag ingeschat.

Use Case P.2: "Als auditor kan ik audio- en videobestanden laten transcriberen, zodat tekstuele inhoud van mediafiles (videobestanden, telefoongesprekken, opgenomen interviews, …) digitaal doorzoekbaar wordt"

Score van de toegevoegde waarde: 8

De toegevoegde waarde van deze functionaliteit wordt als relatief laag ingeschat. De belangrijkste reden hiervoor is dat het gebruik van transcriptietechnologie tijdens interviews de toestemming van de tegenpartij vereist, wat invloed kan hebben op de toon, formaliteit en dynamiek van het interview.

Use Case P.3: "Als auditor kan ik per bestand een samenvattende beschrijving van de inhoud laten genereren, zodat ik snel een indicatie krijg van de inhoud van deze files zonder ze apart te moeten openklikken"

Score van de toegevoegde waarde: 20

Deze *use case* kan een significante meerwaarde realiseren voor het auditproces, met name op het vlak van efficiëntie. Auditoren moeten vaak grote hoeveelheden documenten doorzoeken om relevante informatie te vinden, wat momenteel een tijdrovend en repetitief proces is. Door via een geautomatiseerde manier snel een eerste inschatting te maken van de inhoud, en dus ook de relevantie van documenten, kan deze functionaliteit bijdragen aan een vlottere en meer gestructureerde werkwijze binnen audits.

Use Case P.4: "Als auditor kan ik OCR-processing toepassen op ontvangen bestanden die tekst bevatten, waarbij zeker Nederlands maar eventueel ook Engels, Frans en Duits ondersteund worden, en waarbij de instellingen van de OCR-module aanpasbaar zijn, zodat deze informatie digitaal doorzoekbaar wordt en fouten bij de herkenning van karakters vermeden wordt"

Score van de toegevoegde waarde: 40

Dit is momenteel een basisfunctionaliteit binnen het huidige auditproces; de *use case* wordt dus beoordeeld onder de premisse dat ze een meerwaarde biedt ten opzichte van de huidige tool.

Deze *use case* biedt een duidelijke meerwaarde voor zowel forensische als reguliere audits. Het digitaliseren van ingescande documenten, handgeschreven notities en andere niet-digitale tekstuele bestanden maakt het mogelijk om snel informatie te extraheren uit ongestructureerde data. Dit vergroot de analysemogelijkheden en biedt de kans om transversale zoekopdrachten uit te voeren over de volledige set van bewijsstukken. Momenteel heeft Audit Vlaanderen enkel een OCR-tool ter beschikking binnen het mailboxonderzoek. De meerwaarde van deze *use case* situeert zich daarom op het transversale karakter ervan (zodat zowel documenten als mailboxen volledig digitaal doorzocht kunnen worden).

***Use Case P.5**: "Als auditor kan ik gestructureerde data uit ongestructureerde (via "scraping"/extractie) en gestructureerde bronnen extraheren naar gestandaardiseerde formaten met verwijzing naar de relevante bronbestanden, zodat kwantitatieve informatie uit ongestructureerde bronnen kan doorvloeien naar (al dan niet geautomatiseerde) kwantitatieve analyses zonder manuele extractie en/of preparatie"*

Score van de toegevoegde waarde: **40**

Het overbodig maken van manuele data-extractie en -preparatie leidt tot een significante tijdswinst in het auditproces. Bovendien verlaagt het de kans op fouten, aangezien menselijke tussenkomst in repetitieve taken wordt geminimaliseerd. Volgens de huidige inschatting zijn er voldoende analyses die effectief gestandaardiseerd kunnen worden, waardoor de meerwaarde van deze *use case* als hoog wordt beschouwd.

***Use Case P.6**: "Als auditor kan ik automatisch documenten en e-mails laten dedupliceren (transversaal over alle auditbestanden), zodat niet meer opslagruimte (met bijhorende kost) gebruikt wordt dan nodig"*

Score van de toegevoegde waarde: **13**

Dit is momenteel een basisfunctionaliteit binnen het huidige auditproces; de *use case* wordt dus beoordeeld onder de premisse dat ze een meerwaarde biedt ten opzichte van de huidige tool.

Tijdens de workshop werden meerdere mogelijke meerwaarden aangestipt, waaronder het transversale karakter. Een van de huidige beperkingen binnen de bestaande toolset is dat deduplicatie enkel beschikbaar is voor mailboxen, waardoor dubbele e-mails en bijlagen kunnen worden geïdentificeerd en gefilterd. Voor andere bestandstypes die buiten de mailboxomgeving worden beheerd, ontbreekt deze mogelijkheid. Daarnaast werden ook het goed omgaan met gecompliceerde mailstromen (zoals mailkettingen, personen in CC/BCC, …) en de mogelijkheid om bronbestanden te traceren na deduplicatie als mogelijke meerwaarde vermeld.

***Use Case P.7**: "Als auditor kan ik AI-gegenereerde metadata (inhoud, personen, data, samenvattende beschrijving, …) laten toevoegen aan opgeladen documenten, waarbij steeds duidelijk is waar de metadata op gebaseerd is, zodat mijn zoekopdrachten en analyses ondersteund worden door metadata zonder dat ik deze bestand per bestand zelf moet genereren"*

Score van de toegevoegde waarde: **20**

Deze functionaliteit kan een toegevoegde waarde hebben, omdat ze de auditors toelaat sneller context op te bouwen rond de opgeladen documenten. Er wordt echter wel opgemerkt dat OCR een

belangrijkere functionaliteit vormt (zie *use case* P.4), waarop AI-gegenereerde metadata een goede toevoeging kan zijn.

Use Case E.1: "Als auditor kan ik irrelevante bewijsstukken automatisch laten verbergen (bv. privébestanden, of generieke e-mails zoals nieuwsbrieven of reclame), zodat bewijsstukken die met hoge waarschijnlijkheid irrelevant zijn, mijn zoektocht naar de meest relevante bewijsstukken niet hinderen"

Score van de toegevoegde waarde: **8**

Het automatisch detecteren van (naar hoge waarschijnlijkheid) irrelevante bewijsstukken kan het zoekproces naar relevante stukken efficiënter maken. Er wordt echter aangegeven dat het niet wenselijk is om deze stukken a priori te verbergen, aangezien dit ertoe kan leiden dat informatie die foutief als irrelevant werd bestempeld over het hoofd wordt gezien. In dat opzicht ziet men meer heil in het (al dan niet automatisch) classificeren van de relevantie via metadata (zie *use cases* C.4 en P.7).

Use Case E.2: "Als auditor kan ik een transversale zoekopdracht uitvoeren doorheen alle opgeladen bewijsstukken – op basis van metadata of met exacte/fuzzy zoektermen, waarbij ook synoniemen of gelijkaardige betekenissen in de zoekresultaten opgenomen worden met indicatie van de relevantie(score), zodat bewijsstukken snel gevonden kunnen worden en relevante bewijsstukken niet over het hoofd worden gezien (bv. door verschillende bewoordingen/schrijfwijzen)"

Score van de toegevoegde waarde: **40**

De meerwaarde van deze *use case* wordt zeer hoog ingeschat: de mogelijkheid om alle informatie efficiënt te doorzoeken levert namelijk aanzienlijke tijdswinst op, en verkleint het risico dat cruciale informatie over het hoofd wordt gezien. In de huidige situatie zijn zoekopdrachten niet volledig transversaal, waardoor informatie die verspreid is over verschillende databronnen niet altijd in één zoekactie kan worden doorzocht. Bovendien variëren de zoekmogelijkheden in kracht en precisie, wat ertoe leidt dat bepaalde relevante documenten moeilijker, trager of zelfs niet worden teruggevonden. Een verbeterde zoekfunctionaliteit die alle databronnen integraal en op een krachtige manier doorzoekbaar maakt, zou het auditproces dus aanzienlijk kunnen verbeteren.

Use Case E.3: "Als auditor kan ik matches van zoekresultaten lokaliseren in relevante stukken van documenten of bestanden (bv. sectie/paragraaf bij tekstdocument), zodat ik gericht in dit document kan zoeken naar verdere context"

Score van de toegevoegde waarde: **40**

Deze functionaliteit kan het zoekproces voor auditoren aanzienlijk vergemakkelijken en versnellen. Daarnaast biedt het ook meerwaarde voor het management, aangezien belangrijke informatie tijdens het reviewproces in de rapporteringsfase gemakkelijker teruggevonden kan worden.

Use Case E.4: "Als auditor kan ik suggesties krijgen van contextueel gerelateerde documenten bij het selecteren of openklikken van een bepaald document (bv. document opgesteld in dezelfde tijdsperiode, of vermelding makend van dezelfde personen, …), zodat ik snel verder kan zoeken in documenten met eenzelfde context"

Score van de toegevoegde waarde: **20**

De meerwaarde van deze *use case* wordt zeker erkend, aangezien er tijdens het auditwerk vaak wordt gewisseld tussen verschillende documenten. Dit geldt voor zowel forensische als reguliere audits, hoewel de documentenset bij reguliere audits doorgaans beperkter is. Relevante suggesties kunnen bovendien de blik van de auditor verbreden. Als kanttekening richting implementatie moet echter worden opgemerkt dat een overdaad aan irrelevante of minder relevante suggesties het risico met zich meebrengt dat auditoren het overzicht kunnen verliezen.

***Use Case E.8**: "Als auditor kan ik de organisatie/mappenstructuur van mailboxen en externe opslaglocaties behouden en visualiseren (evt. met weergave per bestandstype/trefwoord en doorklikbare links), zodat relevante onderdelen van mailboxen (bv. project-inboxen) en wanordelijke drives snel geïdentificeerd en geanalyseerd kunnen worden"*

Score van de toegevoegde waarde: **0 (regulier) – 20 (forensisch)**

Voor de forensische auditteams is deze functionaliteit interessant, aangezien het kan bijdragen aan een hogere efficiëntie (door sneller inzicht te verkrijgen in de overvloed aan beschikbare informatie), en aangezien het kan zorgen voor meer transparantie (het is essentieel om duidelijk te maken hoe en waar bepaalde documenten zijn gevonden). Voor de reguliere auditteams is deze functionaliteit echter niet relevant, omdat zij doorgaans niet met e-mails of externe drives in contact komen.

***Use Case E.9**: "Als auditor kan ik de versiegeschiedenis raadplegen van een (gekopieerde) drive/opslaglocatie van een custodian, zodat verwijderde informatie uit eerdere bestandsversies kan opduiken in digitale zoektools"*

Score van de toegevoegde waarde: **3 (regulier) – 20 (forensisch)**

Voor forensische audits is het vergelijken van verschillende versies van groot belang, aangezien wijzigingen in documenten (bijvoorbeeld het weglaten van een bepaalde paragraaf of het herformuleren van een passage) vaak fraude of onregelmatigheden kunnen onthullen. Het zou een meerwaarde bieden als er een loggingfunctionaliteit aanwezig is die automatisch de aanpassingen tussen versies bijhoudt en weergeeft, en als ook verwijderde versies teruggehaald kunnen worden. Voor reguliere audits is het vergelijken van versies doorgaans minder relevant, hoewel de functionaliteit in zeer specifieke gevallen nuttig kan zijn, bijvoorbeeld om te achterhalen hoe vaak een handleiding wordt geüpdatet. Deze *use case* komt inhoudelijk sterk overeen met *use case* I.3 (bestanden inhoudelijk met elkaar vergelijken) en werd daardoor ook hetzelfde gescoord.

***Use Case I.1**: "Als auditor kan ik tekstdocumenten analyseren met network visualization (topical clusters & relationships, word frequency analysis, …), zodat terugkerende thematieken en veelgebruikte bewoordingen duidelijk worden"*

Score van de toegevoegde waarde: **13**

Deze functionaliteit kan het werk van de auditoren versnellen, vooral voor reguliere audits bij lokale besturen, waar het analyseren van tekstdocumenten een groot deel van het werk uitmaakt en tot de belangrijkste activiteiten behoort.

Use Case I.2: "Als auditor kan ik gestructureerde data laten analyseren op patronen, statistische afwijkenen en onlogische combinaties/reeksen, zodat zo veel mogelijk relevante verbanden op efficiënte wijze herkend worden uit een grote set gegevens"

Score van de toegevoegde waarde: **20**

Het analyseren en automatisch detecteren van afwijkingen of verbanden in gestructureerde data kan aanzienlijke efficiëntiewinsten opleveren, vooral bij audits waarbij zeer grote datasets beschikbaar zijn (zoals subsidieaudits). Dit type audit is momenteel voornamelijk van toepassing in het forensische domein, aangezien reguliere audits doorgaans worden gekarakteriseerd door een heterogeen datalandschap met grotere hoeveelheden ongestructureerde data. Aangezien de digitale maturiteit van reguliere audits in de toekomst zal toenemen en er dus steeds meer gestructureerde data geanalyseerd zal worden, wordt deze *use case* voor reguliere audits met dezelfde waarde ingeschat als voor forensische audits, namelijk als een belangrijke groeiopportuniteit.

Use Case I.3: "Als auditor kan ik bestanden inhoudelijk ten opzichte van elkaar laten vergelijken op afwijkingen, zodat verschillen tussen (versies van) bestanden snel gedetecteerd kunnen worden"

Score van de toegevoegde waarde: **3 (regulier) – 20 (forensisch)**

In het kader van huidige forensische audits worden vaak verschillende versies van hetzelfde document met elkaar vergeleken (bijvoorbeeld bij bestekken). Dit kan leiden tot het ontdekken van belangrijke verschillen, zoals een verwijderde paragraaf in een eerdere versie van een document die het ontbrekende puzzelstukje levert voor het blootleggen van verdachte activiteiten. Op dit moment beperkt dit soort onderzoek zich in de praktijk tot een beperkt aantal versies; een dergelijke tool zou een duidelijke meerwaarde kunnen bieden door het mogelijk te maken snel meerdere versies met elkaar te vergelijken. Voor reguliere audits is het vergelijken van bestandsversies minder relevant, waardoor deze *use case* voor hen lager gescoord wordt.

Use Case I.4: "Als auditor kan ik inzichtsvragen stellen aan een AI-chatbot die gevoed wordt met alle verzamelde brondocumenten/bewijsstukken/wetteksten/… , zodat zo veel mogelijk inzichten uit de verzamelde gegevens verzameld worden, inzichten die relevant kunnen zijn om de onderzoeksvragen te beantwoorden"

Score van de toegevoegde waarde: **8 (regulier) – 20 forensisch**

Een AI-chatbot is een zeer vernieuwende functionaliteit die duidelijke meerwaarde kan bieden op het gebied van het genereren van nieuwe inzichten en het leggen van verbanden. Dit is vooral relevant voor forensische audits, waar de omvang van de bewijsstukken doorgaans groot is. Er wordt echter opgemerkt dat men niet altijd op AI-gegenereerde antwoorden kan vertrouwen, waardoor de verantwoordelijkheid voor de interpretatie van de resultaten bij mensen blijft liggen. Als de menselijke verificatie van de AI-antwoorden te veel tijd in beslag neemt, zou deze functionaliteit zelfs het tegenovergestelde effect kunnen hebben van wat beoogd wordt, namelijk de efficiëntie verlagen in plaats van ze te verhogen. Bij de beoordeling van deze *use case* is er een duidelijk verschil merkbaar tussen reguliere en forensische audits, wat te verklaren is door de verschillende scope van beide audittypes: voor forensische audits wordt aan deze functionaliteit een hogere meerwaarde toegekend, omdat daar vaker nieuwe en minder voor de hand liggende denksporen geëxploreerd moeten worden.

***Use Case I.5**: "Als auditor kan ik suggesties krijgen voor ontbrekende documenten/informatie en voor verdere onderzoeksdaden, zodat ik snel bijkomende informatie kan opvragen of nieuwe onderzoekssporen kan exploreren"*

Score van de toegevoegde waarde: **3 (regulier) – 8 (forensisch)**

De waarde van deze *use case* wordt zowel voor reguliere als forensische audits als relatief laag ingeschat. Bij reguliere audits hebben auditoren doorgaans op voorhand een duidelijk beeld van welke documenten en onderzoeksdaden binnen de scope vallen, waardoor ontbrekende stukken snel opgemerkt worden. In forensische audits ligt de situatie anders: hier is het ontbreken van documenten vaak geen toevalligheid, maar een indicatie van mogelijke onregelmatigheden. Dit kan zich uiten in ontbrekende contracten, facturen of correspondentie die normaal gezien aanwezig zouden moeten zijn. Ervaren auditoren herkennen dergelijke hiaten meestal op basis van context en bestaande documentstromen, en weten gericht waar ze verdere informatie kunnen zoeken of aanvullende documentatie kunnen opvragen. Hierdoor wordt de toegevoegde waarde van een automatische detectie van ontbrekende documenten als beperkt beschouwd.

***Use Case I.6**: "Als auditor kan ik (on)gestructureerde gegevens laten verifiëren op basis van verschillende brondocumenten/standaardprocedures, zodat audits verrijkt kunnen worden met gestandaardiseerde tests en illegale activiteiten of afwijkingen op de standaardprocedures sneller (automatisch) gedetecteerd kunnen worden, waardoor de audit sneller in de juiste richting geduwd kan worden"*

Score van de toegevoegde waarde: **20**

Deze *use case* kan zeker een toegevoegde waarde bieden, aangezien (delen van) forensische audits vaak gericht zijn op het onderzoeken van onregelmatigheden binnen specifieke standaardprocedures, zoals subsidiedossiers of overheidsaanbestedingen. Bovendien neemt het handmatig doorzoeken van ongestructureerde gegevens doorgaans veel tijd in beslag. Hulp bij het herkennen van afwijkingen kan dus leiden tot efficiëntiewinsten en ervoor zorgen dat er geen steekproeven nodig zijn binnen de volledige dataset. Dit betekent dat in plaats van enkel een beperkte selectie van gegevens te analyseren, de volledige dataset systematisch kan worden doorgelicht. Hierdoor wordt het risico verkleind dat cruciale onregelmatigheden onopgemerkt blijven, en kunnen analyses met een grotere diepgang en precisie worden uitgevoerd.

***Use Case I.7**: "Als auditor kan ik een process mapping genereren op basis van referenties/gestructureerde data/database logs/etc., waarbij de opeenvolgende stappen bij de opvolging van een dossier alsook de opeenvolgende communicaties en dossierstukken visueel in kaart worden gebracht, en waarbij vergeleken wordt met een "standaardproces", zodat processen bevattelijk kunnen worden voorgesteld en anomalieën tov een typisch procesverloop snel gedetecteerd kunnen worden"*

Score van de toegevoegde waarde: **40**

Dit kan een grote efficiëntiewinst teweegbrengen, aangezien procesanalyse zeer vaak voorkomt in het kader van forensische processen. Bovendien zijn hier momenteel geen toegewijde tools voor voorhanden. Hoewel het creëren van een *process mapping* voorlopig minder aan de orde is voor de

reguliere audits, kan deze functionaliteit voor hen als groeiopportuniteit worden beschouwd om hun huidige toolset uit te breiden en toekomstbestendig te maken (zie ook redenering bij *use case* I.2).

***Use Case I.8**: "Als auditor kan ik een tijdlijn genereren van belangrijke informatie (mailverkeer, dossierbehandeling eventueel met dossierstukken zelf, verklaringen/gebeurtenissen uit interviews, levensloop, …), met doorklikbare links naar de overeenkomstige bronbestanden, zodat belangrijke tijdsinformatie overzichtelijk wordt voorgesteld en ik hiermee inzicht kan verwerven in wat normaal of abnormaal is"*

Score van de toegevoegde waarde: **13 (regulier) – 40 (forensisch)**

Voor forensische audits is deze functionaliteit een absolute gamechanger, omdat een tijdlijn daar vaak hét inzicht oplevert dat een zaak in de juiste richting kan duwen (vaak maakt de timing van een specifieke gebeurtenis het verschil tussen wat als normaal of abnormaal beschouwd wordt). Voor reguliere audits is dit echter minder nuttig, en wordt er meer heil gezien in de functionaliteit rond de *process mapping* (zie *use case* I.7).

***Use Case I.9**: "Als auditor kan ik een visuele voorstelling genereren van communicatiepatronen (wie contacteert wie, met welke frequentie, wanneer, waarover, met welke bijlagen/links), waarbij gefilterd kan worden (op personen/onderwerpen/tijdsintervallen/trefwoorden in bijlagen of mailtekst), zodat opvallende/abnormale communicatiestromen snel gedecteerd kunnen worden"*

Score van de toegevoegde waarde: **1 (regulier) – 20 (forensisch)**

Voor forensische audits kan deze functionaliteit zeker een meerwaarde bieden, aangezien het onderzoek vaak gericht is op het analyseren van relevante communicatiestromen. Het zou bijvoorbeeld nuttig kunnen zijn om spilfiguren te identificeren, en hun communicatie op een overzichtelijke manier in kaart te brengen. Hierbij moet echter rekening worden gehouden met privacykwesties, evenals het feit dat veelvuldige communicatie niet per definitie wijst op verdacht gedrag. Reguliere audits daarentegen richten zich vrijwel nooit op communicatieonderzoek, waardoor deze functionaliteit voor hen aanzienlijk lager wordt gescoord.

***Use Case R.1**: "Als auditor kan ik een doorklikbare onepager laten genereren, zodat de voornaamste conclusies op toegankelijke wijze worden worden gepresenteerd, waardoor ze meer impact kunnen genereren"*

Score van de toegevoegde waarde: **20**

Sommigen schatten de meerwaarde bijzonder hoog in, omdat het helder en toegankelijk presenteren van onderzoeksresultaten kan bijdragen aan een groter bereik en dus ook een grotere impact. Anderen erkennen eveneens de toegevoegde waarde, maar vooral als interne tool in plaats van voor externe verspreiding. Het gebrek aan de nodige nuance maakt deze functionaliteit immers minder geschikt voor externe disseminatie.

***Use Case R.2**: "Als auditor/admin-generaal kan ik finale rapporten automatisch laten klaarmaken voor verzending (incl. toevoegen van watermerk/identificatienummer en uitzwarting indien nodig) en laten verzenden aan de juiste bestemmelingen, zodat ik tijd win door niet naar elke bestemmeling manueel een rapport te moeten uitsturen en niet zelf uitzwarting en identificatiemechanismes te moeten verzorgen"*

Score van de toegevoegde waarde: **0 (regulier) – 13 (forensisch)**

Voor forensische audits ligt de belangrijkste meerwaarde in de automatische uitzwarting, wat een tijdswinst kan opleveren aangezien dit momenteel handmatig gebeurt. De overige functionaliteiten binnen deze *use case*, zoals de toevoeging van identificatiecodes of watermerken, bieden slechts beperkte meerwaarde, aangezien deze mogelijkheden al beschikbaar zijn in het huidige auditproces (in een specifieke verzendmodule). Voor reguliere audits is er op dat vlak geen toegevoegde waarde, aangezien uitzwarting daar niet van toepassing is.

***Use Case R.3**: "Als auditor kan ik op basis van geselecteerde data en tekst een AI-gegenereerde visual toevoegen aan het rapport, bijvoorbeeld van de evolutie van omzettotalen, tijdsverlopen, netwerken van actoren, …, zodat rapporten verrijkt worden met ondersteunende schema's of grafieken"*

Score van de toegevoegde waarde: **13**

De toegevoegde waarde van deze *use case* ligt voornamelijk in het interne gebruik als analysetool, eerder dan als middel om grafieken te genereren voor externe rapportering. Een snel gegenereerde visuele voorstelling van geanalyseerde data kan de interpretatie vergemakkelijken en de inzichten tijdens de analysefase versterken. Het is echter belangrijk op te merken dat de meeste tijd wordt besteed aan de datapreparatie, een proces waarin deze *use case* geen directe meerwaarde biedt.

***Use Case R.4**: "Als auditor kan ik vanuit een vaststellingsdocument of een storyline (in bullets) een voorstel (of set van voorstellen) voor een rapport laten genereren, met correcte referentiëring naar relevante bewijsstukken, zodat er een snelle aanzet is tot rapport die nadien verder aangevuld, gecontroleerd en verfijnd kan worden, maar waarbij er wel steeds waterdichte referentiëring is"*

Score van de toegevoegde waarde: **40**

Deze *use case* wordt als bijzonder waardevol beschouwd, aangezien er aanzienlijke tijdswinst kan worden geboekt in de rapporteringsfase, die momenteel erg arbeidsintensief is. De mogelijkheid om een automatisch gegenereerde eerste versie van een rapport te verkrijgen, maakt het mogelijk sneller over te gaan naar de revisiefase, waar de inhoud verder wordt geverifieerd, uitgewerkt en verfijnd.

***Use Case R.5**: "Als auditor kan ik (rapport)teksten laten herschrijven (zij het naar een consistente stijl/formaat, zij het naar een bondigere synthese/samenvatting), zodat alle rapporten (voor disseminatie naar verschillende kanalen) voldoen aan verwachtingen in termen van kwaliteit (to-the-point, hoge leesbaarheid en consistente stijl)"*

Score van de toegevoegde waarde: **100**

Deze *use case* wordt beschouwd als een absolute gamechanger voor zowel reguliere als forensische audits. De onderliggende redenering sluit aan bij die van *use case* R.4. Deze functionaliteit biedt de mogelijkheid om de rapporteringsfase, die momenteel bijzonder tijdsintensief is, aanzienlijk te versnellen. Daarnaast draagt ze bij aan een consistente schrijfstijl, wat de kwaliteit en professionaliteit van de rapportering versterkt. Net als bij *use case* R.1 kan dit ook het bereik en de impact verhogen.

***Use Case R.6**: "Als auditor kan ik automatisch documenten (bv. rapport, audit brief / opdrachtdefinitie, …) laten aanvullen met identificatie-informatie (bv. auditcode, …), zodat alle documenten gerelateerd aan eenzelfde audit op betrouwbare wijze geïdentificeerd kunnen worden"*

Score van de toegevoegde waarde: **besproken maar niet gescoord (meegenomen in R.2)**

***Use Case R.7**: "Als auditor kan ik autocorrectie toepassen tijdens het typen van namen, zodat doorheen de rapportering consistente naamgeving gebruikt wordt"*

Score van de toegevoegde waarde: **5 (regulier) – 13 (forensisch)**

Deze functionaliteit kan een nuttige aanvulling zijn op de toolset binnen de rapporteringsfase, aangezien consistentie een belangrijke indicator is van professionaliteit. Voor reguliere audits wordt de meerwaarde als beperkter beschouwd dan bij forensische audits, aangezien in reguliere audits minder individuele namen worden vermeld en er vaker in algemene termen naar departementen of entiteiten wordt verwezen (bv. de HR-dienst of de financiële cel in plaats van specifieke medewerkersnamen), waardoor het risico op fouten kleiner is.

***Use Case R.8**: "Als auditor kan ik tijdens de rapportering namen van custodians, organisaties, etc. selecteren uit een lijst, zodat er een 'single source of truth' is voor naamgeving, en doorheen de rapportering consistente naamgeving gebruikt wordt"*

Score van de toegevoegde waarde: **niet besproken**

***Use Case M.1**: "Als manager-auditor kan ik een overzicht raadplegen van de auditstatus (lijst met onderzoeksvragen, lijst van personen waarmee gesproken werd, overzicht van analyses/documenten/… die door mensen gemaakt dan wel door AI gegenereerd werden, …), zodat dit centraal overzicht mij ondersteunt bij het bepalen van welke audits mijn aandacht nodig hebben, en zodat ik steeds op de hoogte ben van alle stappen die al gezet zijn"*

Score van de toegevoegde waarde: **20**

Deze functionaliteit kan van waarde zijn voor manager-auditors om het overzicht over lopende audits te behouden en de planning effectief te bewaken. Vandaag hanteert elke manager-auditor doorgaans zijn eigen methode om dit op te volgen. De meerwaarde van deze functionaliteit zou er dan ook in bestaan om de opvolging van audits op uniforme wijze te organiseren binnen één gedeelde tool, wat kan bijdragen tot een meer gestructureerde en coherente sturing.

***Use Case M.2**: "Als auditor kan ik straf- en tuchtrechterlijke gevolgen van afgelopen audits opvolgen, zodat ik kan inschatten of verdere follow-up noodzakelijk is"*

Score van de toegevoegde waarde: **0 (regulier) – 3 (forensisch)**

Voor forensische audits worden momenteel eenmaal per jaar gegevens opgevraagd bij de geauditeerde entiteit of de politiediensten. Gezien deze beperkte frequentie is de toegevoegde waarde van deze functionaliteit eerder laag. Voor reguliere audits is het opvolgen van straf- en tuchtrechterlijke gevolgen niet van toepassing.

***Use Case M.3**: "Als auditor kan ik alle bewaarde data (met versiegeschiedenis) van afgelopen audits raadplegen in een database, zodat eerder opgevraagde/verzamelde informatie die thans relevant is, kan worden hergebruikt (minstens als eerste check, alvorens de actualiteit ervan te checken)"*

Score van de toegevoegde waarde: **13 (regulier) – 5 (forensisch)**

Er kan een duidelijk onderscheid worden gemaakt tussen reguliere en forensische audits. Bij reguliere audits kan dit een meerwaarde bieden voor de procesbeheersing, in de zin dat auditoren beter zicht krijgen op de manier waarop een organisatie haar processen structureert, opvolgt en bijstuurt. Door eerdere audits van dezelfde entiteit te raadplegen, kunnen vastgestelde risico's opnieuw onder de loep worden genomen en kan worden nagegaan welke aandachtspunten inmiddels zijn aangepakt en welke nog verdere opvolging vereisen. Daarnaast kan het nuttig zijn om bij de start van een audit context op te bouwen, door documenten zoals het organigram, de deontologische code en het arbeidsreglement te raadplegen. Voor forensische audits is dit minder relevant, aangezien eenzelfde entiteit zelden meermaals wordt geauditeerd. Bovendien zijn gegevens met betrekking tot forensische audits vaak uiterst gevoelig. Daarom is het essentieel om afdoende beveiligings- en beschermingsmechanismen te voorzien voor de opslag en toegankelijkheid van deze data.

2.2.4 Randvoorwaarden

De voorgestelde IT-tooling moet aan een aantal minimumvereisten of (technische) randvoorwaarden voldoen. Deze randvoorwaarden worden hieronder opgesomd.

<u>Gegevensbescherming</u>

Audit Vlaanderen heeft een decretaal mandaat om bij audits de nodige gegevens te verzamelen, ongeacht de drager, ook gegevens met een hogere gevoeligheid en/of vertrouwelijkheid. IT-tooling ter ondersteuning van de auditprocessen moet daarom ook de nodige maatregelen op vlak van gegevensbescherming kunnen bieden.

<u>Bestandsvalidatie en beveiliging</u>

De tool moet bij het uploaden van bestanden automatische *sanity checks* uitvoeren, zoals virusscans, volledigheidscontroles en integriteitschecks. Dit zorgt voor een veilige en goed beheerde opslagomgeving, en maakt snelle detectie en reactie op problematische bestanden mogelijk.

<u>Metadata</u>

Auditoren moeten metadata kunnen toevoegen en aanpassen om bewijsstukken efficiënt te categoriseren, labelen en annoteren. Deze metadata moeten vervolgens doorzoekbaar zijn en ondersteuning bieden bij het onderbouwen van vaststellingen, conclusies en verantwoordingen binnen het auditproces. Bovendien moet het systeem filtermogelijkheden bieden op basis van de gedefinieerde metadata, zodat de weergegeven zoekresultaten steeds relevant zijn.

<u>Bestandsverwerking en integriteit</u>

De tool moet verwerking van bestanden mogelijk maken binnen een gecontroleerde werkomgeving waarbij de oorspronkelijke bronbestanden onaangetast blijven. Dit voorkomt juridische betwistingen over de authenticiteit van de data en biedt een betrouwbare basis voor verdere analyses.

<u>Ondersteuning van diverse bestandsformaten</u>

De IT-tooling moet een breed scala aan gestructureerde en ongestructureerde bestandstypes kunnen verwerken, waaronder Office-bestanden, PDF's, PST-mailbestanden, multimediaformaten en XML. Dit garandeert dat relevante bewijsstukken veelal ongehinderd kunnen worden gebruikt in de auditcontext.

Veilig delen van bestanden

Auditoren moeten op een veilige manier bestanden of subsets daarvan kunnen delen, bijvoorbeeld voor gedetailleerde analyses die uitbesteed worden aan externe partners. Dit vereist gecontroleerde toegangsrechten. Ook binnen de eigen organisatie moet het mogelijk zijn om bestanden en analyses eenvoudig te delen met collega's, zodat samenwerking binnen het auditteam soepel verloopt en informatie-uitwisseling efficiënt gebeurt.

3 Analyse van de "state of the art"

3.1 Referentiearchitecturen

Op basis van de geanalyseerde noden binnen het auditproces (zie Sectie 2) werden twee referentiearchitecturen geïdentificeerd. Deze referentiearchitecturen verschillen in de dekking van de functionele noden en in de invulling van het innovatieve karakter. Ze zijn schematisch weergegeven in Figuur 4 en Figuur 5.

De functionele noden zijn geclusterd in groene bouwstenen, die elk een specifieke processtap ondersteunen. Hieronder volgt per processtap een beknopte toelichting van de bijhorende functies:

- Collect:
  - **Upload (meta)data**: een uploadmodule waarmee de geauditeerde entiteit eigen data kan aanleveren, al dan niet verrijkt met optionele metadata;
  - **Follow-up collection**: een opvolgingsmodule voor het systematisch inventariseren van gevraagde en ontvangen bestanden of informatie;
  - **Sanity checks**: automatische kwaliteits- en veiligheidscontroles op binnenkomende bestanden, zoals virusscans, bestandsvalidatie en integriteitscontrole;
- Process:
  - **Optical Character Recognition (OCR)**: omzetting van tekst in afbeelding naar een machine-leesbaar digitaal formaat;
  - **Transcription**: transcriptiemodule voor de omzetting van audio naar tekst;
  - **Metadata extraction**: module voor automatische extractie van (technische) metadata uit de verzamelde bestanden;
  - **Custodian linking**: het koppelen van bestanden, e-mails of andere documenten aan betrokken personen binnen het auditdossier;
  - **Deduplication**: het opsporen en verwijderen van dubbele bestanden of e-mails;
  - **Transformation**: omzetting van bestanden naar een gestandaardiseerd formaat om verdere verwerking mogelijk te maken;
  - **Categorisation**: tagging/labeling van verzamelde bestanden voor verdere verwerking;
  - **Batch rename**: functionaliteit om meerdere bestanden gelijktijdig te hernoemen volgens een vast naamgevingsconventie;
  - **Folder & file management**: beheer van de digitale opslagstructuur, inclusief opschoning en ordening van bestanden;
- Explore/insight:

  - **Search**: transversale zoekmodule die het mogelijk maakt de verzamelde set van informatie efficiënt te doorzoeken;
  - **Generic visualisation**: generieke visualisaties, zoals grafieken of dashboards, die inzicht geven in kwantitatieve gegevens;
  - **Specific visualisation**: audit-specifieke visualisaties zoals communicatiepatronen, tijdslijnen, …;
  - **Standard analyses**: klassieke statistische analysemethoden op gestructureerde data, zoals regressieanalyses of detectie van uitschieters;
  - **AI analyses**: geavanceerde data-analyses op basis van AI/ML-technieken, zoals *(un)supervised learning*, AI-gebaseerde *outlier detection* en NLP-gebaseerde tekstanalyse;
  - **Virtual assistant**: een virtuele assistent waarmee de auditor kan interageren om taken uit te voeren of de beschikbare informatie te doorzoeken;
- Report:
  - **AI (re)write**: generatieve tekstmodule voor het schrijven van een rapport(aanzet), of het herschrijven van bestaande rapporten/verslagen met een consistente stijl en formaat;
  - **Submit module**: functionaliteit voor het voorbereiden en versturen van eindrapporten, inclusief het toekennen van identificatiecodes en routing naar de juiste bestemmeling;
  - **Auto censoring**: module voor het automatisch uitzwarten van gevoelige informatie (bv. persoonsgegevens) bij de disseminatie van eindverslagen.

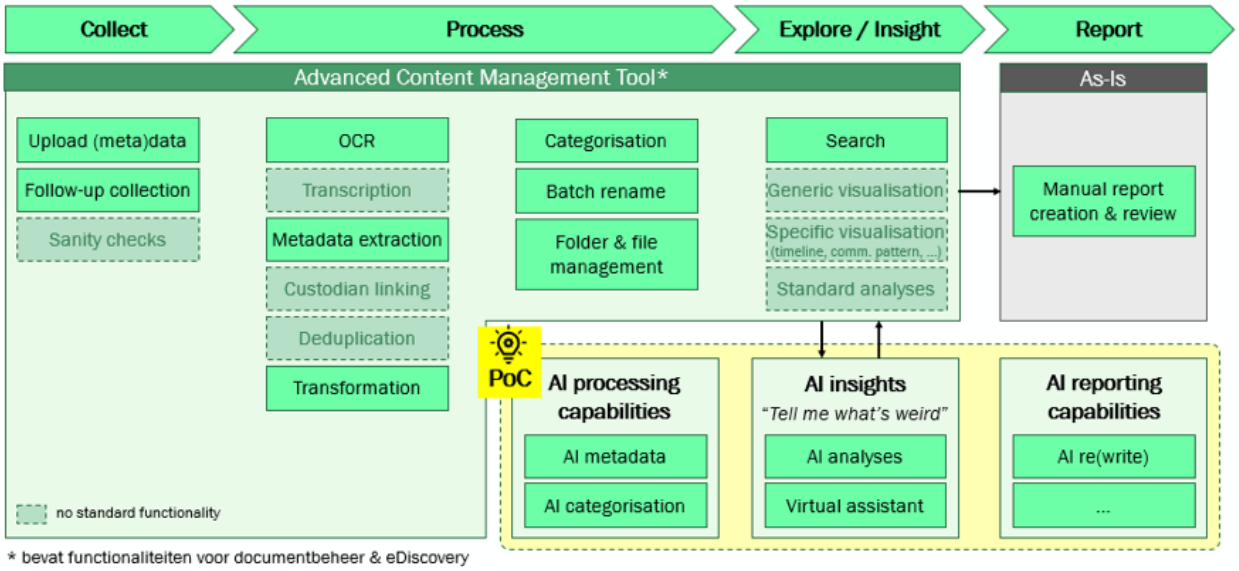
Referentiearchitectuur 1 is schematisch weergegeven in Figuur 4Figuur 4: Referentiearchitectuur 1 ("Content manager met innovatieve AI-component").. Deze architectuur vertrekt van het uitgangspunt dat een aanzienlijk deel van de functionele bouwstenen binnen het auditproces kan worden ingevuld met behulp van een geavanceerde contentmanagement tool. Deze functionaliteiten situeren zich voornamelijk in de beginfasen van het auditproces – met name *Collect*, *Process*, *Explore* en *Insight* – en richten zich in essentie op generiek documentbeheer, eventueel aangevuld met *e-discovery*-capaciteiten.

De lichtgroene blokken in Figuur 4 vertegenwoordigen functionaliteiten die doorgaans standaard aanwezig zijn in dergelijke contentmanagementsystemen. Deze omvatten onder meer modules voor documentenupload en -opvolging, *optical character recognition* (OCR), metadata-extractie, datatransformatie, categorisatie en zoekfunctionaliteiten. De donkergroene blokken met stippellijnkaders verwijzen naar functionaliteiten die niet standaard zijn, maar die afhankelijk van de gekozen softwareoplossing — en mits eventuele integraties of maatwerk — wel gerealiseerd kunnen worden. Dit betreft onder meer transcriptie, deduplicatie, geavanceerde analyses en datavisualisatie.

Het innovatieve element van deze referentiearchitectuur schuilt in de integratie van een veelzijdige AI-component, die als *piloot-traject* wordt opgezet ter aanvulling van het contentmanagement-platform. Deze AI-module kan op uiteenlopende manieren bijdragen aan een efficiënter auditproces: van intelligente documentclassificatie en automatische extractie van relevante informatie, tot het genereren van (deel)analyses of het ondersteunen bij het schrijven en herwerken van rapportteksten.

Aangezien de architectuur zich hoofdzakelijk toespitst op de eerste fasen van het auditproces, wordt de grootste tijdswinst verwacht bij het verzamelen, verwerken en verkennen van auditinformatie. De

bestaande werkwijze voor de rapporteringsfase – die grotendeels manueel blijft en meerdere reviewmomenten omvat – blijft in deze architectuur behouden.



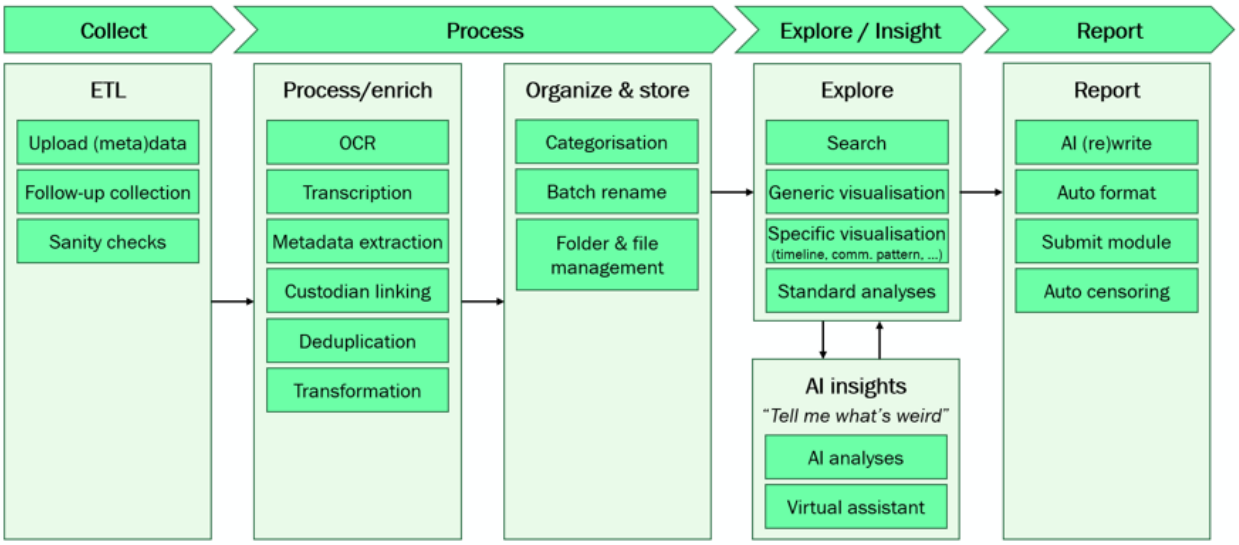
Figuur 4: Referentiearchitectuur 1 ("Content manager met innovatieve AI-component").

Referentiearchitectuur 2, weergegeven in Figuur 5, onderscheidt zich van Referentiearchitectuur 1 door haar bredere scope: waar die laatste zich richt op specifieke stappen binnen het auditproces, biedt deze architectuur een oplossing die alle functionele behoeften van begin tot eind afdekt — de zogenaamde 'full package'.

De architectuur is opgebouwd uit een reeks functionele bouwstenen waarbinnen de functionele noden, zoals hiervoor vermeld, werden ondergebracht:

- een *Extract/Transform/Load* (ETL)-module voor het verzamelen en structureren van gegevens,
- een verwerkingsmodule voor het voorbereiden en verrijken van de verzamelde data,
- een organisatiemodule voor de gestructureerde opslag en categorisering van informatie,
- een exploratiemodule met zoek-, analyse- en visualisatiefuncties,
- een AI-module met zowel virtuele assistentie als AI-gedreven inzichten, en
- een rapporteringsmodule voor het opstellen en finaliseren van (eind)verslagen.

De technische invulling van deze modules kan op verschillende manieren gebeuren: via een combinatie van afzonderlijke, gespecialiseerde tools, of via de implementatie van een meeromvattende softwaresuite. In beide gevallen kunnen bijkomend maatwerk of specifieke integraties vereist zijn, afhankelijk van de specifieke behoeften en beschikbare functionaliteiten. De innovatieve waarde van deze architectuur is dus sterk afhankelijk van de gekozen implementatiestrategie. Door haar focus op een zo volledig mogelijke functionele dekking, levert deze referentiearchitectuur tijdswinst op in alle stappen van het auditproces.



Figuur 5: Referentiearchitectuur 2 ("full package").

3.2 Marktanalyse

In de marktanalyse werden bestaande softwareoplossingen gezocht die de bovenstaande behoeften deels of volledig opvangen. Op basis van de functionele behoeften werden er zes categorieën van tools geïdentificeerd:

1. **'Dedicated audit tools'**: deze tools zijn specifiek gericht op auditprocessen, en bieden functionaliteiten die zich uitstrekken over de verschillende processtappen (*Collect*, *Process*, *Explore*, *Insight Generation*, *Report* en *Management*). In tegenstelling tot generieke analysetools, zijn 'dedicated audit tools' doorgaans geconfigureerd volgens auditstandaarden en bevatten ze ingebouwde controles om het werk van auditors te vergemakkelijken. Er zijn verschillende tools die zich op andere types audits richten: forensische audits, financiële audits, interne compliance audits, ...
2. **'Content management tools'**: deze tools worden ingezet voor het beheren, opslaan en organiseren van digitale content, zoals documenten, e-mails en mediabestanden. Ze richten zich voornamelijk op het efficiënt indexeren en doorzoekbaar maken van informatie, en bieden daarnaast functies zoals versiebeheer, toegangscontrole en metadata-ondersteuning.
3. **'Data analytics tools'**: deze tools worden gebruikt voor het analyseren en interpreteren van gestructureerde gegevens om patronen, trends en afwijkingen te identificeren. Binnen het auditdomein stellen deze tools auditors in staat om snel grote en complexe datasets te doorzoeken. Naast zuivere analysefunctionaliteiten bieden sommige van deze tools ook functies aan om data uit verschillende bronnen te consolideren en te transformeren naar gewenste (standaard)formaten.
4. **'Data visualisation tools'**: deze tools bieden de mogelijkheid om gegevens visueel weer te geven in de vorm van grafieken, diagrammen, dashboards of andere interactieve visualisaties. Dit helpt om efficiënt inzichten te verwerven op basis van grote en complexe datasets. Deze tools versterken niet alleen de interne dataverkenning tijdens de audit zelf, maar zijn ook waardevol voor het helder presenteren van resultaten of bevindingen aan stakeholders.

5. **'Text mining/analysis tools'**: deze tools zijn gericht op het verwerken, analyseren en interpreteren van ongestructureerde tekstdata, zoals documenten, e-mails, verslagen, notulen of transcripties van interviews. Auditors kunnen deze tools inzetten om snel relevante informatie, veelvoorkomende bewoordingen, thematieken, sentimenten of risico's te identificeren uit grote volumes van tekstuele bronnen (waarvoor de manuele analyse zeer tijdrovend zou zijn).
6. **'(Generative) AI tools'**: deze tools maken gebruik van kunstmatige intelligentie om complexe taken te ondersteunen of autonoom uit te voeren, zoals het genereren van tekst, samenvattingen, inzichten of aanbevelingen. Generatieve AI-tools vormen een subcategorie binnen de bredere AI-tooling: deze tools zijn in staat om op basis van bestaande data of een gegeven context nieuwe inhoud te genereren.

In totaal werden 43 commercieel beschikbare softwareoplossingen geëvalueerd op de mate waarin ze de functionele noden van Audit Vlaanderen (zie Sectie 2.2.3) ondersteunen. De details van deze analyse zijn opgenomen in Annex A. Hieronder worden de belangrijkste inzichten opgesomd:

- Er bestaan zowel gespecialiseerde tools voor specifieke doeleinden (zoals data-analyse of *text mining*) als brede softwarepakketten (suites) die meerdere auditfunctionaliteiten integreren.
- Geen enkele van de onderzochte 'dedicated audit tools' dekt alle functionele noden van Audit Vlaanderen volledig af. Voor een oplossing die wél alle behoeften ondersteunt (zoals in Referentiearchitectuur 2, zie Figuur 5) zijn er dus twee benaderingen mogelijk:
    1. Het combineren van meerdere gespecialiseerde *best-in-class* tools. Deze aanpak biedt inhoudelijke sterkte, maar leidt tot een niet-homogene en dure oplossing (bv. door het opstapelen van meerdere licentiekosten).
    2. Het inzetten van een uitgebreide softwaresuite. Dergelijke pakketten bieden meer samenhang en gebruiksgemak, maar behoren tot de duurdere opties, en nemen in kostprijs toe naarmate er meer modules moeten worden toegevoegd. Daartegenover staat dat veel *best-in-class* functionaliteiten al zijn ingebed in grotere suites.
- Als niet alle functionele noden moeten worden afgedekt, kan een geavanceerd contentmanagementsysteem een haalbaar alternatief zijn (zoals in Referentiearchitectuur 1, zie Figuur 4). Uit de marktanalyse blijkt dat de meest mature contentmanagementtools een groot deel van de functionele behoeften kunnen invullen (met name binnen de processtappen *Collect*, *Process* en *Explore*). Hier is ruimte voor een hybride oplossing op maat, waarbij de innovatieve AI-component kan worden gerealiseerd met de ondersteuning van een derde partij die als (IT-)integrator optreedt.

4 Marktconsultatie

Op basis van de analyse van de state of the art (zie Sectie 3) werd een strategische richting uitgewerkt bestaande uit 3 luiken:

1. **Normale overheidsopdracht**: om de continuïteit te waarborgen voor de huidige werking (gegeven de aflopende raamcontracten voor de huidige *e-discovery*-software), wordt binnen het eerste luik een reguliere overheidsopdracht uitgeschreven. Het doel van deze

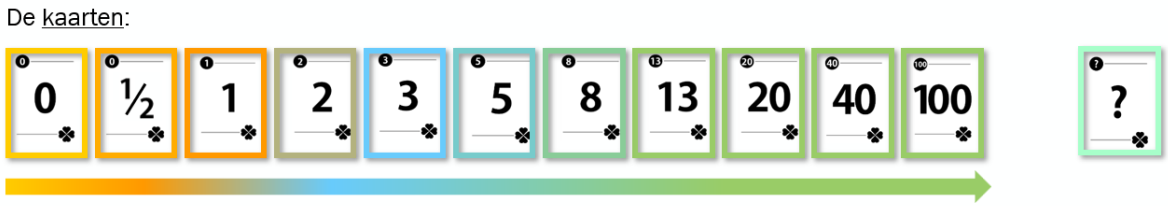
overheidsopdracht is om op korte termijn de huidige manier van werken te kunnen voortzetten, zodat de innovatieve overheidsopdracht in voldoende detail kan worden voorbereid.

2. ***AI-driven piloot-traject* op korte termijn**: aangezien het innovatieve karakter van de referentiearchitecturen en de state-of-the-art tooling zich op het gebied van artificiële intelligentie bevindt, kan men een piloot-project uitwerken voor de aanpak van enkele *high-value* uitdagingen op korte termijn. In deze piloot wordt gefocust op enkele afgelijnde *use cases* waar AI-technologie centraal staat.
3. **Scope-uitbreiding op lange(re) termijn**: de piloot uit het vorige luik kan nadien worden uitgebreid naar andere *use cases* en overheidsdiensten; deze uitbreiding focust zich meer op de lange termijn en benut opportuniteiten die voortvloeien uit de snelle technologische evolutie waar AI vandaag door wordt gekenmerkt.

Meer details over deze aanpak worden vermeld in Sectie 5.

Aangezien dit voortraject als doel heeft om een innovatieve aanbesteding voor te bereiden (zie Sectie 1.2.3), werd gefocust op luik 2 wegens het sterke innovatieve karakter. Ook de bevraging van de markt werd vanuit dit uitgangspunt georganiseerd: uit de originele *use cases* (zie Sectie 2.2.3) werd een selectie gemaakt van 24 *use cases* die hierbinnen passen, en deze *use cases* werden vervolgens geherformuleerd met meer focus op een AI-component.

De geherformuleerde *use cases* werden tijdens de plenaire sessie van de marktconsultatie voorgelegd aan het panel van ingeschreven experten, en beoordeeld op technische haalbaarheid. Dit werd opnieuw gedaan via de *planning poker* techniek (zie Sectie 2.2.3). Belangrijk om op te merken is dat de *use cases* werden beoordeeld met een kortetermijnperspectief ("wat is de haalbaarheid **op dit moment**?"). Onderstaande figuur geeft een overzicht van de mogelijke scores en hun interpretaties:



De interpretatie voor de spelers:

0-2	**Off-the-shelf**: beschikbaar binnen een redelijk geprijsd, flexibel en schaalbaar commercieel product.
3-5	**Minimal effort**: commercieel product beschikbaar, minimale aanpassingen nodig & eenvoudige integratie.
8-13	**Mogelijk**: er is een redelijke commerciële basis, maar enige aanpassing/integratie/expertise vereist.
20-40	**Uitdagend**: zwakke commerciële basis, vereist sterke aanpassingen en complexe integratie.
100	**Onmogelijk!** Vereist een custom design over meerdere expertisedomeinen en/of tart de wetten van de fysica
?	**Geen idee**, geen ervaring met dit onderwerp.

Hieronder worden de consensusscores voor de *use cases* opgelijst, inclusief een synthese van de belangrijkste elementen ter motivering. Voor de eenvoud wordt, naast de score voor de technische complexiteit, ook de eerder bepaalde score voor de toegevoegde waarde (van de originele *use case*) herhaald. In onderstaande lijst werden zowel de originele nummering van de *use cases* als de originele prefixen behouden (zie Sectie 2.2.3); om duidelijk aan te geven dat deze *use cases* geherformuleerd werden naar een meer AI-gerichte oplossing werd een asterisk (*) toegevoegd.

Use Case P.3: "Als auditor kan ik generatieve AI gebruiken om automatisch per bestand een samenvattende beschrijving van de inhoud te laten genereren, zodat ik snel een indicatie krijg van de inhoud zonder elk bestand afzonderlijk te moeten openklikken"*

Score van de toegevoegde waarde: **20**
Score van de technische complexiteit: **5**

Voor het genereren van samenvattende beschrijvingen van tekstbestanden bestaan reeds tal van tools en technologieën die deze usecase ondersteunen. Wanneer er ondersteuning vereist is voor een bredere waaier aan documenten (zowel qua bestandstype als qua inhoud), neemt de complexiteit enigszins toe. Deze complexiteit ligt niet zozeer in de technische implementatie zelf, maar eerder in de zogenaamde *prompt engineering* — het verfijnen van de input om de gewenste output te verkrijgen (voor elk document).

Use Case P.7: "Als auditor kan ik AI-gegenereerde metadata (inhoud, personen, data, samenvattende beschrijving, …) laten toevoegen aan opgeladen documenten, waarbij steeds duidelijk is waar metadata op gebaseerd is, zodat mijn zoekopdrachten en analyses ondersteund worden door metadata zonder dat ik deze bestand per bestand zelf moet genereren"*

Score van de toegevoegde waarde: **20**
Score van de technische complexiteit: **13**

Op technologisch vlak zijn er zowel commerciële oplossingen als initiatieven in ontwikkeling binnen (Belgische) onderzoeksinstellingen. Het mappen van informatie op een gestandaardiseerde ontologie met een specifiek datamodel vormt op zich geen grote technologische uitdaging. De moeilijkheid zit veeleer in het scherp definiëren van de vereisten: hoe het datamodel eruit moet zien en welke metadata per bestandstype geëxtraheerd moeten worden. Hoewel deze usecase technisch uitvoerbaar is, moet wel rekening gehouden worden met vereisten rond de opslag van gevoelige data. Dit kan de complexiteit verhogen, aangezien de oplossing dan *on-premises* moet worden uitgerold, zonder data-uitwisseling met de publieke cloud. Dit laatste is een transversaal inzicht, dat van toepassing is op meerdere *use cases*.

Use Case P.5: "Als auditor kan ik AI gebruiken om gestructureerde data automatisch te extraheren uit ongestructureerde bronnen (via scraping of extractie) en gestructureerde gegevensbronnen, en deze om te zetten naar gestandaardiseerde formaten met verwijzing naar de relevante bronbestanden, zodat kwantitatieve informatie uit ongestructureerde bronnen direct kan doorvloeien naar geautomatiseerde analyses zonder manuele extractie of voorbereiding"*

Score van de toegevoegde waarde: **40**
Score van de technische complexiteit: **8**

Dataextractie uit gestructureerde bronnen is een veelvoorkomend probleem dat geen echte intrinsieke technologische struikelblokken kent. De invulling ervan is wel sterk afhankelijk van wat je precies wil extraheren (bv. dossiernummers, leveranciers, datums, etc.), en van wat de kwaliteit is van de brondocumenten. Indien de brondocumenten een standaardopmaak of -structuur bevatten is het een eenvoudiger probleem dan wanneer er geen consistente lay-out of plaatsing van informatie is.

Use Case E.4: "Als auditor kan ik AI inzetten om automatisch contextueel gerelateerde documenten aan te bevelen bij het selecteren of openen van een bepaald document (bv. document opgesteld in*

dezelfde tijdsperiode, of vermelding makend van dezelfde personen, …), zodat ik snel verder kan zoeken in documenten met dezelfde context"

Score van de toegevoegde waarde: **20**
Score van de technische complexiteit: **13**

Technologie voor het clusteren of indexeren van documenten op basis van voorgedefinieerde metadata is reeds commercieel beschikbaar. Wanneer er echter ook met inhoudelijke/contextuele elementen rekening moet worden gehouden verhoogt de complexiteit vrij snel, en schaalt ze met de omvang van de gecollecteerde gegevens: er moet een vectordatabase over alle documenten worden opgebouwd, en vaak is specifieke training vereist om de zoek- of suggestiefunctionaliteit optimaal te laten functioneren. Bovendien moeten er controlemethoden worden ingebouwd om te verzekeren dat relevante documenten niet over het hoofd worden gezien; dit moet voor elke audit gebeuren.

Use Case E.1*: "Als auditor kan ik ML- en NLP-technologieën gebruiken om automatisch irrelevante items (bv. privébestanden of generieke e-mails zoals nieuwsbrieven of reclame) te identificeren en te verbergen, zodat mijn zoektocht naar relevante bewijsstukken niet gehinderd wordt door onbelangrijke documenten/mails"*

Score van de toegevoegde waarde: **8**
Score van de technische complexiteit: **5**

Deze *use case* is technologisch niet complex en valt onder de categorie van classificatieproblemen. Ze is vergelijkbaar met het principe van een *junk folder* bij e-maildiensten, waarbij irrelevante of ongewenste informatie automatisch gefilterd wordt. De uitdaging zit enkel in het scherp definiëren van de criteria voor irrelevantie of ongewenstheid, en in het trainen van het systeem om foutieve beslissingen zoveel mogelijk te vermijden.

Use Case I.7*: "Als auditor kan ik AI gebruiken om automatisch een proces mapping te genereren op basis van referenties/gestructureerde data/database logs/etc., waarbij de opeenvolgende stappen, communicatie en dossierstukken in een dossier visueel worden weergegeven en vergeleken met een "standaardproces", zodat processen bevattelijk kunnen worden voorgesteld en anomalieën t.o.v. een typisch procesverloop efficiënt gedetecteerd kunnen worden"*

Score van de toegevoegde waarde: **40**
Score van de technische complexiteit: **40**

Deze *use case* wordt als technisch uitdagender beschouwd, voornamelijk door de variëteit aan referenties en de noodzaak tot een heldere definitie van het 'standaardproces'. Wanneer men bijvoorbeeld een intern proces binnen een gecontroleerde entiteit wil analyseren, is het essentieel dat dit proces gedetailleerd en correct is gedocumenteerd. Zoniet is een accurate vergelijking niet mogelijk. Visuele weergaves – zoals schematische voorstellingen van het procesverloop – vallen bovendien momenteel buiten de mogelijkheden van de meest gangbare AI-systemen. Op dit laatste vlak is er wel een sterke technologische voortuitgang merkbaar; toekomstperspectieven op basis van deze (verwachte) evolutie werden bewust niet meegenomen in de huidige toegekende score(s).

Use Case I.6*: "Als auditor kan ik met behulp van AI (on)gestructureerde gegevens laten verifiëren op basis van verschillende brondocumenten/standaardprocedures, zodat audits verrijkt kunnen worden met gestandaardiseerde tests en zodat illegale activiteiten of afwijkingen op de standaardafspraken*

automatisch gedetecteerd kunnen worden (waardoor de audit sneller in de juiste richting geduwd kan worden)"

Score van de toegevoegde waarde: **20**
Score van de technische complexiteit: **20**

Voor deze *use case* wordt er vanuit gegaan dat de brondocumenten/standaardprocedures die als referentie dienen voldoende gedetailleerd en accuraat gedocumenteerd zijn. De combinatie van verschillende types referentiedocumenten (wetteksten, tabellen, tekstbestanden, etc.) vereist wel specifieke domeinkennis en mogelijk de integratie van meerdere AI-componenten. Bovendien moet de AI-component in staat zijn om te redeneren op een incomplete set van informatie, die niet altijd rechtlijnig interpreteerbaar is. In het bijzonder bij juridische teksten kan het nodig zijn om de juridische basis door een expert te laten vertalen naar een meer AI-vriendelijk formaat.

Use Case I.8*: *"Als auditor kan ik AI inzetten om automatisch een tijdlijn te genereren van belangrijke informatie (mailverkeer, dossierbehandeling met bijbehorende dossierstukken, verklaringen uit interviews, etc.) met doorklikbare links naar de bijbehorende brondbestanden, zodat tijdsgebonden informatie overzichtelijk wordt voorgesteld en ik snel inzicht krijg in wat (ab)normaal is"*

Score van de toegevoegde waarde: **13 (regulier) – 40 (forensisch)**
Score van de technische complexiteit: **13**

Er zijn reeds verschillende tools die deze functionaliteit kunnen aanbieden op basis van brondocumenten die aangevuld zijn met de nodige (*timestamp*-gebaseerde) metadata. De complexiteit wordt groter wanneer, zoals bij Audit Vlaanderen, tijdsgebonden informatie uit de context of uit de inhoud van bestanden geëxtraheerd moet worden, of wanneer de dataset bijzonder omvangrijk is. In dat geval kan het een uitdaging zijn om het juiste detailniveau te bepalen.

Use Case I.2*: *"Als auditor kan ik AI gebruiken om gestructureerde data automatisch te laten analyseren op patronen, statistische afwijkingen en onlogische combinaties, zodat relevante verbanden of anomalieën efficiënt herkend kunnen worden binnen grote datasets"*

Score van de toegevoegde waarde: **20**
Score van de technische complexiteit: **5**

De analyse van (grote hoeveelheden) structurele data behoort tot de klassieke problemen binnen de *machine learning* (ML), waarvoor er verschillende kant-en-klare oplossingen beschikbaar zijn. Niettemin blijft het belangrijk om de geschikte technologie te kiezen (statistische methodes versus getrainde AI-modellen), en dient er steeds een vorm van manuele interactie met of input van de auditor voorzien te worden.

Use Case I.4*: *"Als auditor kan ik inzichtsvragen stellen aan een AI-chatbot die gevoed wordt met alle verzamelde brondocumenten/bewijsstukken/wetteksten/… , zodat zo veel mogelijk inzichten uit de verzamelde gegevens verzameld worden, inzichten die relevant kunnen zijn om de onderzoeksvragen te beantwoorden"*

Score van de toegevoegde waarde: **8 (regulier) – 20 forensisch**
Score van de technische complexiteit: **3**

Het ontwikkelen en trainen van AI-chatbots om inputbestanden conversatiegewijs te doorzoeken vormt geen grote uitdaging en is al veelvuldig toegepast, ook in *on-premise* omgevingen (bv. met *rack*-architecturen). Het is echter wel zo dat er een sterke nadruk ligt op het ingeven van de correcte prompts. Daarnaast kan het auditspecifieke karakter – waarbij men eerder per audit een aparte chatbot wil dan één overkoepelend systeem – bijkomende complexiteit met zich meebrengen, om vermenging van gegevens of biases te vermijden.

***Use Case I.9**: "Als auditor kan ik met behulp van AI een visuele voorstelling genereren van communicatiepatronen (wie contacteert wie, met welke frequentie, wanneer, waarover, met welke bijlagen/links), waarbij machine learning-modellen gefilterde analyses uitvoeren op personen/onderwerpen/tijdsintervallen/trefwoorden in bijlagen of mailtekst), zodat abnormale of opvallende communicatiestromen snel gedetecteerd kunnen worden"*

Score van de toegevoegde waarde: **1 (regulier) – 20 (forensisch)**
Score van de technische complexiteit: **13**

Vanuit technisch oogpunt wordt deze *use case* als matig tot goed haalbaar ingeschat. Enerzijds vertoont ze gelijkenissen met bestaande *e-discovery*-functionaliteiten of *process mining*-technieken. Anderzijds ligt de moeilijkheid in het bepalen van wat als abnormaal gedrag moet worden beschouwd: frequente communicatie tussen individuen is immers niet per definitie verdacht. Bovendien blijft het automatisch genereren van visuele representaties op dit vlak nog steeds een uitdaging voor veel AI-modellen.

***Use Case I.3**: "Als auditor kan ik gebruikmaken van NLP-technologie om bestanden inhoudelijk met elkaar te vergelijken, zodat semantische/structurele verschillen tussen (versies van) documenten automatisch gedetecteerd kunnen worden"*

Score van de toegevoegde waarde: **3 (regulier) – 20 (forensisch)**

Score van de technische complexiteit: **5**

Structurele verschillen tussen documenten – zoals het ontbreken van een paragraaf in een recentere versie – kunnen eenvoudig worden opgespoord via standaardtechnologie. Interpretatie van inhoudelijke of contextuele verschillen vormen daarentegen een grotere uitdaging. Een hybride oplossing, waarbij een deterministisch systeem verantwoordelijk is voor de layoutvergelijking en een LLM de contextuele verschillen detecteert, kan hier uitkomst bieden. Ook in deze *use case* speelt vertrouwelijkheid een belangrijke rol, waardoor cloudgebaseerde oplossingen mogelijk geen optie zijn.

***Use Case I.1**: "Als auditor kan ik AI-gedreven tekstanalyse inzetten die thematische clusters, semantische relaties en woordfrequenties als netwerkvisualisatie toont, zodat ik sneller inzicht krijg in terugkerende onderwerpen en opvallende taalpatronen"*

Score van de toegevoegde waarde: **13**
Score van de technische complexiteit: **3**

Aangezien het hier veeleer om thematische analyse gaat is dit een *use case* die eenvoudig kan worden opgevangen met bestaande technologie: methodes zoals *topic modeling*, *clustering* en trendanalyse zijn reeds veelvuldig toegepast in uiteenlopende domeinen.

Use Case I.5: "Als auditor kan ik AI gebruiken die via patroonherkenning en kennismodellen suggesties geeft voor ontbrekende documenten of aanvullende onderzoeksdaden, zodat ik snel ontbrekende informatie kan opvragen of nieuwe onderzoekssporen kan exploreren"*

Score van de toegevoegde waarde: 3 (regulier) – 8 (forensisch)
Score van de technische complexiteit: 20

Deze *use case* is iets complexer, vooral wanneer het AI-systeem autonoom moet aangeven welke informatie mogelijk ontbreekt. Het voorspellen van ontbrekende elementen zonder expliciete verwijzingen is immers bijzonder moeilijk. Wanneer er wel aanwijzingen of verwijzingen naar ontbrekende documenten aanwezig zijn, wordt de analyse aanzienlijk haalbaarder. Men moet ook rekening houden met de proportionaliteit van de oplossing: de voorgestelde denksporen of onderzoeksdaden moeten relevant zijn en voldoende onderbouwing bevatten.

Use Case R.5: "Als auditor kan ik met behulp van generatieve AI (rapport)teksten laten herschrijven (zij het naar een consistente stijl/formaat, zij het naar een bondigere synthese/samenvatting), zodat alle rapporten (voor disseminatie naar verschillende kanalen) voldoen aan verwachtingen in termen van kwaliteit (to-the-point, hoge leesbaarheid en consistente stijl)"*

Score van de toegevoegde waarde: 100
Score van de technische complexiteit: 3

Generatieve AI voor het (her)schrijven van teksten is vandaag een zeer mature technologie. Wel stijgen de complexiteit en de bijhorende kosten aanzienlijk indien verwerking via de cloud uitgesloten is vanwege vertrouwelijkheid. Voor het formatteren van teksten naar een specifieke layout zijn bovendien ook traditionele, niet-AI-gebaseerde oplossingen beschikbaar.

Use Case R.4: "Als auditor kan ik generatieve AI gebruiken om vanuit een vaststellingsdocument of een storyline in bullets automatisch een voorstel of set van voorstellen voor een rapport te laten genereren, met correcte referentiëring naar relevante bewijsstukken, zodat ik snel een onderbouwde en waterdichte basisversie van het eindrapport krijg die verder kan worden aangevuld, gecontroleerd of verfijnd"*

Score van de toegevoegde waarde: 40
Score van de technische complexiteit: 5

De argumentatie voor deze *use case* loopt vrij gelijk met die voor R.5*. Wel vormt de vereiste dat rapportage foutloos en juridisch sluitend moet zijn een bijkomende uitdaging: LLM's zijn per definitie nooit volledig waterdicht, en er zal dus steeds menselijke controle of verificatie nodig blijven.

Use Case R.1: "Als auditor kan ik generatieve AI gebruiken om automatisch een doorklikbare onepager te laten genereren die de voornaamste conclusies visueel en tekstueel samenvat, zodat bevindingen op een toegankelijke en impactvolle manier gepresenteerd kunnen worden"*

Score van de toegevoegde waarde: 20
Score van de technische complexiteit: 13

De complexiteit hiervan hangt sterk af van de gewenste output. Het genereren van een eenvoudige HTML-pagina met doorklikbare hyperlinks is geen moeilijke opdracht, maar het 'from scratch' genereren van een AI-visual met toevoeging van tekstuele data is een stuk complexer. Dit laatste geldt

zeker aangezien op dit moment, zoals ook vermeld bij enkele andere usecases, de beschikbare AI-technologie nog niet matuur genoeg is op vlak van visualisaties. Gezien de razendsnelle evolutie van de technologie rond afbeeldingsgeneratoren kan men echter verwachten dat ook deze *use case* in de toekomst meer binnen handbereik komt. Intussen kan al tijdswinst geboekt worden door AI suggesties te laten doen over hoe informatie het best kan worden weergegeven, ook al genereert het systeem de visuals nog niet zelf.

Use Case R.3: "Als auditor kan ik op basis van geselecteerde data en tekst een AI-gegenereerde visual toevoegen aan het rapport, bijvoorbeeld van de evolutie van omzettotalen, tijdsverlopen, netwerken van actoren, …, zodat rapporten verrijkt worden met ondersteunende schema's of grafieken"*

Score van de toegevoegde waarde: **13**
Score van de technische complexiteit: **13**

De argumentatie voor deze *use case* loopt vrij gelijk met die voor R.1*. Het genereren van visuals met correcte tekstuele of numerieke annotaties blijft uitdagend maar is niet onmogelijk. Er bestaan ook al out-of-the-box oplossingen die dashboards weergeven (bv. van omzettotalen), maar hier is de assumptie dat de nodige metadata beschikbaar is.

Use Case R.2: "Als auditor/admin-generaal kan ik AI met NLP/NER gebruiken om finale rapporten automatisch klaar te maken voor verzending (waarbij gevoelige informatie automatisch wordt uitgezwart), zodat ik tijd win en de vertrouwelijkheid gegarandeerd blijft zonder manuele tussenkomst"*

Score van de toegevoegde waarde: **0 (regulier) – 13 (forensisch)**
Score van de technische complexiteit: **13**

Het anonimiseren of pseudonimiseren van namen is technisch relatief eenvoudig bij consistente naamgeving. De moeilijkheidsgraad stijgt echter wanneer bescherming tegen zowel directe als indirecte identificatie vereist is – bijvoorbeeld wanneer de identiteit afgeleid kan worden uit de omschrijving, zonder dat de naam expliciet genoemd wordt – of wanneer dezelfde persoon of entiteit op verschillende manieren benoemd wordt (initialen, pseudoniemen, afkortingen, alternatieve schrijfwijzen, etc.). Bijkomend geldt dat dergelijke functionaliteit weliswaar beschikbaar is bij verschillende aanbieders, maar dat een volledige *on-premises*-implementatie vanwege vertrouwelijkheid aanzienlijke kosten met zich meebrengt, aangezien dergelijke verwerking vaak in de publieke cloud plaatsvindt.

Use Case C.8: "Als auditor kan ik generatieve AI gebruiken om op basis van (informatie uit) vorige audits een passend juridisch advies te formuleren, zodat adviezen niet dubbel opgevraagd moeten worden informatie uit vorige audits efficiënt hergebruikt kan worden in een andere context"*

Score van de toegevoegde waarde: **20**
Score van de technische complexiteit: **8**

Op voorwaarde dat de brondata van juridische adviezen correct en actueel is, vormt deze *use case* technisch geen grote uitdaging. De functionaliteit komt overeen met die van een chatbot die vragen beantwoordt op basis van een vooraf gedefinieerde set documenten (zie usecase I.4*). De uitdaging zit hier minder in het AI-stuk zelf, maar eerder in het bijhouden en valideren van de brondata in een

contentmanagementsysteem. Vorige adviezen kunnen namelijk hun geldigheid verliezen wanneer de juridische basis ervoor gewijzigd wordt.

Use Case C.6: "Als auditor kan ik AI inzetten om (bijvoorbeeld op basis van inhoudsanalyse via OCR en NLP) automatisch bestandsnamen toe te kennen aan ontvangen PDF's/foto's/etc., zodat onduidelijke bestandsnamen het auditproces niet vertragen en zodat manuele filenaming niet meer vereist is"*

Score van de toegevoegde waarde: **13**
Score van de technische complexiteit: **5**

Deze *use case* is technisch goed haalbaar en vereist een combinatie van OCR (voor afbeeldingen, bv. foto's van documenten) en NLP-technologie (voor machine-leesbare tekstbestanden). Beide technologieën zijn inmiddels commercieel ingeburgerd en breed beschikbaar.

Use Case C.3: "Als auditor kan ik AI gebruiken om automatisch suggesties te krijgen voor op te vragen (standaard)documenten op basis van een opgegeven auditcategorie, waarbij een ML-model patronen in eerdere audits analyseert en vaakvoorkomende documenttypes suggereert, zodat deze vaakvoorkomende documenttypes niet over het hoofd worden gezien bij het opstellen van de inventaris van op te vragen documenten"*

Score van de toegevoegde waarde: **5**
Score van de technische complexiteit: **5**

Deze *use case* is technisch zeer haalbaar: een LLM kan auditcategorieën analyseren op basis van beschikbare data uit vorige audits, waarna een statistisch model kan bepalen welke documenttypes het vaakst voorkomen binnen een bepaalde categorie.

Use Case M.1: "Als manager-auditor kan ik AI gebruiken om, binnen een systeem dat de status van lopende audits bijhoudt, automatisch suggesties te krijgen over welke audits mijn aandacht vereisen, zodat potentiële risico's binnen lopende audits snel gedetecteerd kunnen worden"*

Score van de toegevoegde waarde: **20**
Score van de technische complexiteit: **8**

Voor deze *use case* kan bestaande projectmanagementsoftware een solide basis bieden. Deze kan eventueel worden aangevuld met een AI-component die op basis van deterministische indicatoren – zoals doorlooptijd of het aantal openstaande vragen – mogelijke knelpunten detecteert. De implementatie van dergelijke functionaliteit is technisch haalbaar, mits integratie met een bestaande logica en de beschikbaarheid van de juiste indicatoren.

5 Synthese van het verdere traject

5.1 High-level vormgeving van aanbestedingsstrategie

De voorgestelde richting van het vervolgtraject (reeds kort aangehaald in Sectie 4) bestaat uit drie luiken, die zowel inspelen op de onmiddellijke nood aan continuïteit als op de opportuniteiten die AI

op korte en lange termijn biedt. Elk luik vervult een specifieke rol in het gefaseerd realiseren van een toekomstbestendige digitale oplossing voor Audit Vlaanderen.

Luik 1: Normale overheidsopdracht voor continuïteit op de korte termijn

Gezien het aflopen van de huidige raamcontracten voor de *e-discovery*-software, is het essentieel om de bestaande operationele continuïteit van Audit Vlaanderen te garanderen. In het eerste luik wordt daarom een gewone overheidsopdracht uitgeschreven, met als doel de huidige manier van werken te kunnen verderzetten met minimale aanpassingen aan het originele procesverloop.

De focus van deze aanbesteding ligt op een kortetermijnoplossing die voldoende aansluit bij de bestaande *toolset* en werkmethodiek, zonder hierbij nieuwe en ingrijpende technologiekeuzes te introduceren. Deze tussenstap schenkt AV voldoende tijd om de innovatieve componenten van de toekomstige aanbesteding(en) zorgvuldig vorm te geven.

Luik 2: *AI-driven piloot-traject* op korte termijn

In het tweede luik wordt een experimentele en toekomstgerichte aanpak voorgesteld: het opzetten van een beperkte maar doelgerichte piloot rond AI. De technologische evolutie en de combinatie van de referentiearchitecturen met de analyse van de state of the art (zie Sectie 3) tonen immers aan dat net binnen het AI-domein de meest vernieuwende mogelijkheden liggen voor het automatiseren en versnellen van complexe taken binnen het auditproces. Dit argument wordt versterkt door de observatie dat de auditaanpak van Audit Vlaanderen weliswaar binnen een gestandaardiseerd procesverloop kadert (zie Sectie 2.2.2), maar telkens voldoende case-specifiek is zodat klassieke automatiseringsmethoden (zoals RPA) minder geschikt zijn.

Het voorgestelde piloot-traject richt zich dan ook op enkele scherp afgebakende *use cases* met hoge meerwaarde. Door gericht te experimenteren met bestaande AI-tools en -technieken kan op korte termijn worden geëvalueerd welke technologieën effectief geschikt zijn voor grootschalige inzet binnen de (forensische of reguliere) auditcontext. Deze piloot biedt niet alleen de mogelijkheid tot verkenning van de technologische mogelijkheden, maar ook een kans om gebruikservaringen in de praktijk te verzamelen, technische risico's of struikelblokken te identificeren, en op die manier waardevolle input te leveren voor de uitwerking van een bredere strategische roadmap.

Luik 3: Op lange(re) termijn: uitbreiding van de scope en opschaling

Het derde luik bouwt voort op de inzichten en resultaten van de piloot in het tweede luik, en richt zich op een gefaseerde uitbreiding van de reikwijdte en impact van de AI-oplossing. Deze fase beoogt een structurele verankering van innovatieve technologie binnen Audit Vlaanderen (en eventueel ook binnen andere overheidsdepartementen).

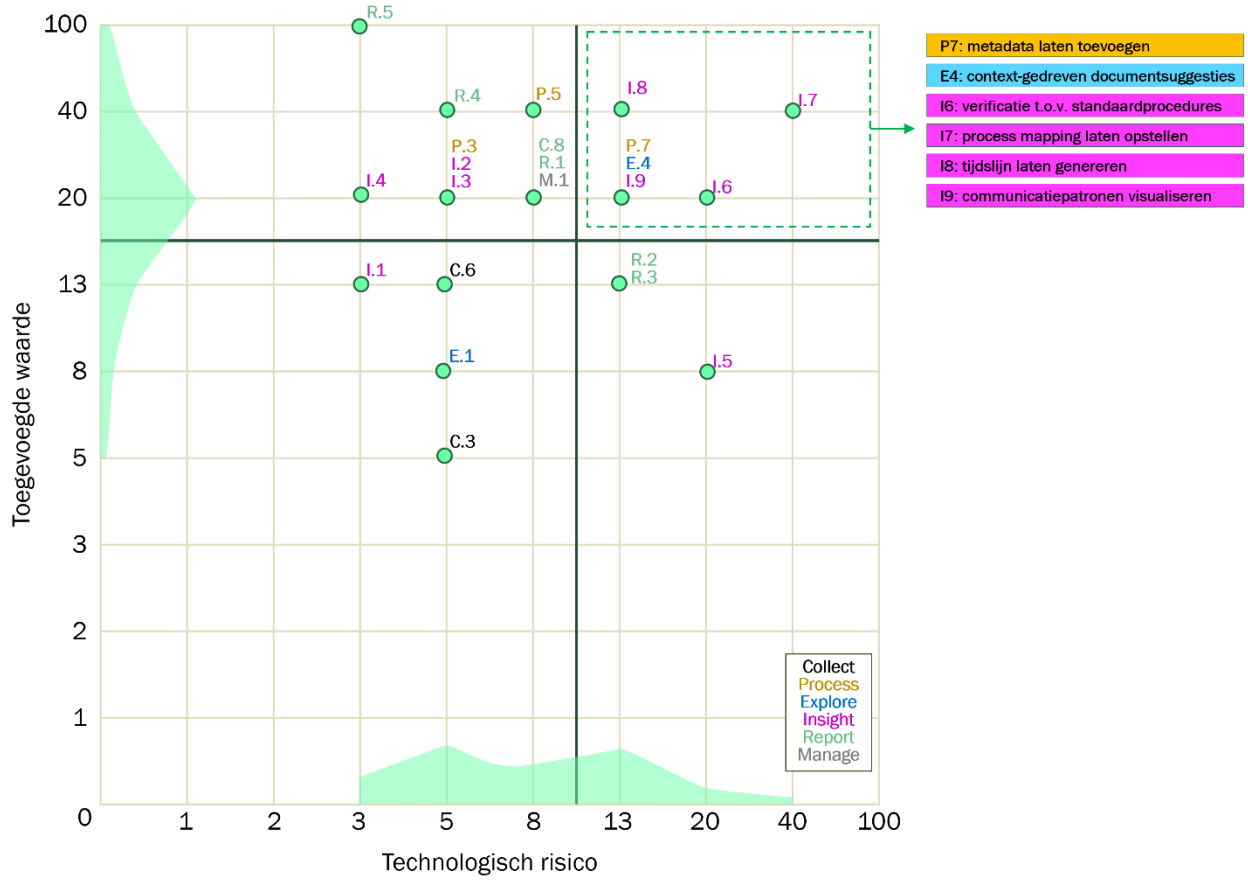
De scope-uitbreiding van verschillende vormen aannemen. De meest voor de hand liggende vorm is het opschalen naar bijkomende *use cases* — diegenen die reeds geïdentificeerd waren, of nog aan de oppervlakte komen binnen het tweede luik. Ook synergieën met andere

auditfuncties binnen overheidsdiensten of de academische sector behoren tot de mogelijkheden.

Belangrijk aan dit luik is dat het voldoende flexibiliteit behoudt om in te spelen op de razendsnel evoluerende AI-markt. Door modulair en incrementeel te werk te gaan, wordt vermeden dat technologiekeuzes verouderen nog voor de volledige uitrol heeft plaatsgevonden.

5.2 Inschatting van het innovatiepotentieel

Het innovatiepotentieel van de *use cases* kan worden ingeschat op basis van een waarde-risico matrix, weergegeven in onderstaande figuur. Op de verticale as bevindt zich het innovatiepotentieel vanuit het gebruikersperspectief, m.a.w. de toegevoegde waarde die voor elke *use case* werd ingeschat door middel van een *planning poker* sessie met de auditor en auditor-managers (waarvan het resultaat is neergeschreven in Sectie 2.2.3). De horizontale as geeft het technologisch risico van de *use cases* weer, verkregen door een *planning poker* sessie met spelers uit de industrie tijdens de marktconsultatie. De geconsolideerde scores met achterliggende motivatie achter deze scores zijn gedocumenteerd in Sectie 4.



Op de verticale as, dus volgens het aspect 'toegevoegde waarde', zien we een sterke clustering tussen de waardes van 13 t.e.m. 40. Dit is niet onverwacht, aangezien deze subset van *use cases* net werd

geselecteerd als waardevolle kandidaten voor een kleine innovatieve aanbesteding rond AI-gebaseerde functionaliteiten. Op de horizontale as, dus volgens het aspect 'technologisch risico', zien we een sterkere spreiding (voornamelijk tussen 3 en 20).

Use cases E.1 (rond het automatisch verbergen van irrelevante informatie), C.3 (rond automatische suggesties voor op te vragen documenten) en I.5 (rond automatische suggesties voor bijkomend onderzoek) werden op vlak van toegevoegde waarde laag ingeschat, wat het weinig interessant maakt om hierin te investeren. Buiten deze drie *use cases* liggen de inschattingen voor toegevoegde waarde over het algemeen hoog: alle overige *use cases* scoren 13 of hoger op de verticale as.

Richting een innovatieve aanbesteding gaan we op zoek naar waardevolle en uitdagende *use cases*. We delen het vlak daarom op in 4 kwadranten, als eerste indicatie van de interessantste *use cases*. Op beide assen leggen we de scheidingslijn halverwege de waargenomen verdelingen (en daarom iets verder op de verticale as dan op de horizontale as).

In het kwadrant rechtsboven bevinden zich zo 6 *use cases* met sterk innovatiepotentieel: zij kunnen grote toegevoegde waarde leveren voor het auditproces, en kunnen ondersteund worden mits de nodige integraties en ontwikkelingen op maat gebeuren. Daarbij kan men wel vertrekken van bestaande commerciële producten, of producten die al in ontwikkeling zijn.

Meerdere marktspelers gaven de feedback dat de *use cases* scherp geformuleerd waren, maar in hun geheel een heel brede scope omvatten. De gedetailleerde *use cases* zijn een pluspunt richting het uitwerken van een aanbesteding, de brede scope vormt een risico richting de implementatie (omwille van hogere totale complexiteit en verspreide focus). Om ook de scope scherp te houden, stellen wij vast dat meer dan de helft van de *use cases* in het kwadrant rechtsboven voortkomen uit de 'Insight generation'-stap van het audit proces.

Het feit dat het technologisch risico (van het merendeel van de *use cases*) afhangt van de mate waarin de betrokken gegevens beschermd moeten worden, vormde een tweede terugkerend aandachtspunt. Neem bv. *use case* R.2, rond het automatisch uitzwarten van gevoelige informatie: deze veronderstelt dat de AI-systemen wel toegang hebben tot de gevoelige informatie die verborgen moet worden (naar bepaalde lezers van rapporten toe). Neem als tegenvoorbeeld *use case* R.3, rond het laten genereren van visuals die de rapporten ondersteunen: dit soort *use cases* zou in principe kunnen werken met enkel de informatie die ook in de eindrapporten beschikbaar blijft. De keuze tussen een *on-premise*, *private cloud* of *public cloud* architectuur ligt dus gevoeliger bij *use cases* zoals R.2 dan bij *use cases* zoals R.3.

Verscheidene marktspelers gaven ook aan dat bij langdurige AI-implementatietrajecten, bv. wanneer men de scope niet goed bewaakt of wanneer men vooral ambitie eerder dan ervaring heeft, de resultaten soms al achterhaald zijn nog voor zij worden gerealiseerd. AI-technologie evolueert snel, dus voor partijen zoals AV, met weinig AI-ervaring, loont het de moeite om in te zetten op tastbare, puntsgewijze verbeteringen. Een POC kan typisch binnen enkele weken opgeleverd worden, maar hierdoor wordt de positieve impact op de dagdagelijkse werking nog niet gerealiseerd. Het doorduwen tot een volwaardige piloot-implementatie (in productie) vergt doorgaans nog eens 2 tot 3 keer zo veel tijd.

5.3 Conclusie

Uit de bovenstaande risico-inschattingen en de feedback uit de marktconsultatie blijkt dat heel wat functionaliteiten die AV voor ogen heeft, kunnen worden ondersteund, vertrekkende van bestaande AI-gedreven producten en AI-gedreven producten in ontwikkeling. Daar waar de meeste proces-stappen kunnen verrijkt worden door (quasi) afgewerkte producten, biedt vooral de 'Insight-generation'-stap veel innovatiepotentieel met hoge toegevoegde waarde voor de auditors en auditor-managers. De uitdagingen van dit project liggen op de volgende vlakken:

- Keuze van architectuur
  AV kan (zelfstandig) kiezen voor een architectuur die het beste past bij de *use cases* die zij willen prioriteren, maar de vraagstukken rond gegevensbescherming en cloudmigraties leven ook op transversaal niveau binnen de Vlaamse Overheid. Uiteraard moeten afwijkingen ten opzichte van richtlijnen op transversaal niveau vermeden worden, maar op dit moment zijn niet alle richtlijnen reeds beschikbaar. De gefaseerde aanpak die AV beoogt (eerst een klein piloot-project en daarna scope-uitbreiding) verschaft enige tijd waarin deze richtlijnen verder vorm kunnen krijgen, maar om verdere aanbestedingen niet in de weg te staan kan AV best enige flexibiliteit op vlak van architectuur reeds inbouwen in de eerste aanbesteding.

- Tijdelijk karakter van auditprocessen (en de gevarieerde onderliggende datasets)
  Elke doorloop van het auditproces gebeurt in het algemeen op basis van een geheel nieuwe set van gegevens. Enkele van de *use cases* hebben wel betrekking tot gegevens die blijven bestaan buiten de context van individuele audits, zoals *use case* C.8 (rond het hergebruiken van juridische adviezen), maar zelfs daar blijft het idee om deze persistente gegevens te verwerken binnen de context van individuele audits (dus steeds samen met auditspecifieke gegevens). Dit betekent dat een aantal stappen per audit herhaald moet worden, waaronder stappen die normaliter (bij andere toepassingen) niet herhaald hoeven te worden (bv. pre-processing of training). Om verrassingen (op vlak van kosten) te vermijden, zal het belangrijk zijn voor AV om hun noden hieromtrent te specifiëren in een aanbesteding, en voor kandidaten om de bijhorende kosten hieromtrent exhaustief te benoemen.

- Snelle evolutie van AI-technologie
  Het wordt AV aanbevolen om erover te waken dat al haar inspanningen rond AI-gedreven procesverbeteringen snel tot tastbare resultaten komen, die ook effectief hun ingang vinden in de dagelijkse werking van de auditors. Zo kan er zelfs voor een eerste aanbesteding een volwaardige business case worden opgesteld. De 'lessons learned' uit zo'n aanbesteding zullen dan ook meer gewicht dragen en betere inzichten voortbrengen richting toekomstige grootschaligere aanbestedingen rond AI.

De gewonnen inzichten wijzen in de richting van een gefaseerde aanbesteding, zoals eerder besproken: een klassieke overheidsopdracht om de continuïteit te garanderen, vervolgens een kleinschalig AI-gedreven pilootproject, en daarna stelselmatige scope-uitbreidingen. Bovendien kunnen synergiën met andere auditfuncties op termijn ook zorgen voor een grotere afzetmarkt van de hierin ontwikkelde oplossingen.

6 Bronnen

[1] https://www.vlaanderen.be/organisaties/administratieve-diensten-van-de-vlaamse-overheid/beleidsdomein-kanselarij-bestuur-buitenlandse-zaken-en-justitie/audit-vlaanderen

[2] https://www.auditvlaanderen.be/over-audit-vlaanderen

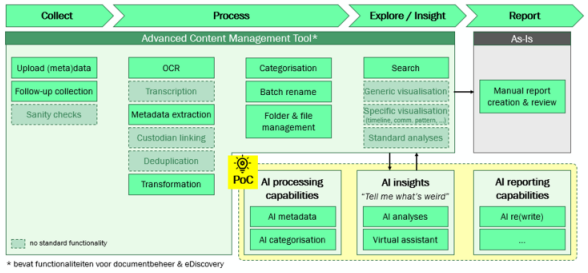
[3] https://www.auditvlaanderen.be/over-audit-vlaanderen/soorten-audits

[4] https://www.vlaio.be/nl/vlaio-netwerk/programma-innovatieve-overheidsopdrachten-pio

Annex A: Marktanalyse

We zien 2 referentiearchitecturen met verschillende dekking van de noden

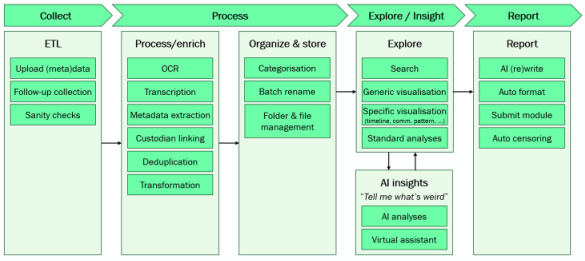
Referentiearchitectuur 1:
Content manager* met AI



- Focus op Collect, Process & Explore
- Innovatieve component zit in veelzijdige AI-integratie
- Kan ingevuld worden met geavanceerde content management tool, al dan niet met custom development of extra integraties

Conclusie
- Tussenliggende oplossing met innovatieve component
- Grootste (tijds)winst in process/explore/insight

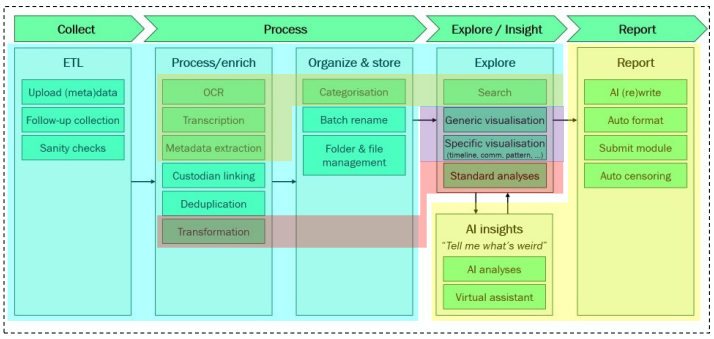
Referentiearchitectuur 2:
"Full package"



- Dekt alle functionele noden van A tot Z
- Innovatieve component afhankelijk van keuze voor implementatie
- Kan zowel ingevuld worden door combinatie van afzonderlijke tools als door een complete suite (met custom development)

Conclusie
- Duurste maar meest complete oplossing
- (Tijds)winst in alle stappen van het auditproces

We gingen op zoek naar bestaande oplossingen
op basis van de 6 categorieën die de process-stappen deels of volledig opvangen



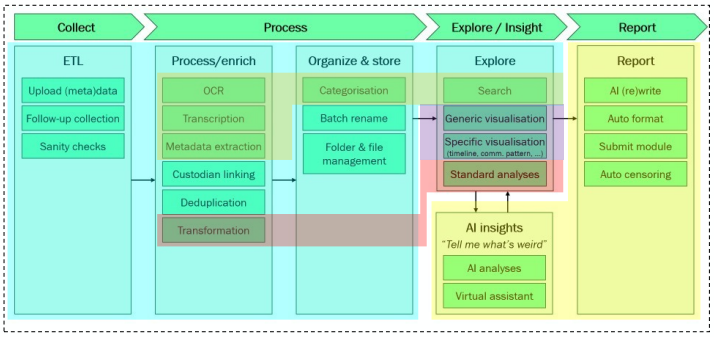
5 categorieën voor *specifieke functionaliteiten*

1. Content management
2. Data analytics
3. Data visualisation
4. Text mining & analysis
5. (Generative) AI

+ 1 overkoepelende categorie

6. Dedicated audit tool

In totaal werden 43 tools geëvalueerd over deze 6 categorieën

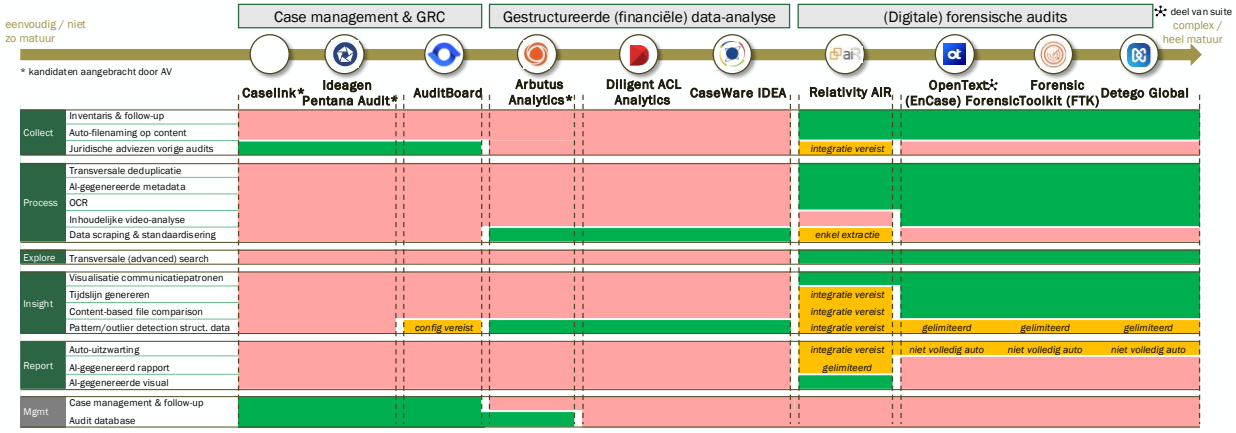


	5 catogorieën voor *specifieke functionaliteiten*	# tools
1	Content management	3
2	Data analytics	11
3	Data visualisation	6
4	Text mining & analysis	7
5	(Generative) AI	6
	+ 1 overkoepelende categorie	
6	Dedicated audit tool	10

Dedicated Audit Tool

Beschrijving

- ✓ Ingebouwde workflows & analysemodules speciaal gericht op auditprocessen (forensisch, GRC, ...)
- ✓ (Intelligente) gegevensverwerking met import-, analyse- & visualisatiefunctionaliteiten
- ✓ Efficiënte documentatie, referentiëring & gestandaardiseerde rapportering



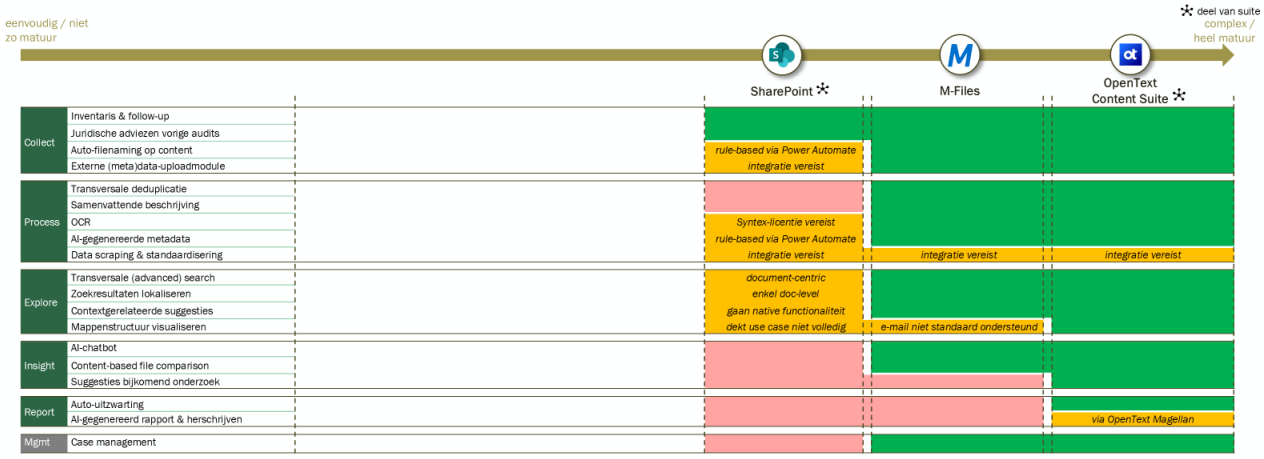
Voornaamste takeaways

- ➢ **Geen enkele tool kan alles** maar mits toegevingen zijn er wel sterke kandidaten
  - ➢ Sommige tools gericht op compliance (bv. AuditBoard), andere op financiële fraude (bv. Caseware IDEA) of eDiscovery (bv. FTK)

Content Management Tool

Beschrijving

- ✓ Centrale opslag & organisatie voor documenten, media en andere content
- ✓ Zoekfuncties, tagging/categorisering & ondersteuning voor metadata
- ✓ Versiebeheer en samenwerking via toegangsrechten



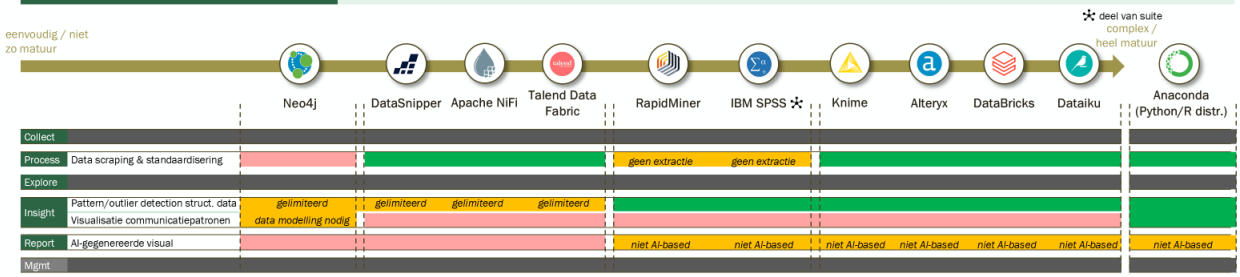
Voornaamste take-aways

- ➢ Geavanceerde content management tools dekken veel use cases over alle stappen van het auditproces
- ➢ Verschil zit vooral in maturiteit van de zoekfuncties, AI-functionaliteiten & integratiemogelijkheden (bv. SharePoint in Microsoft 365-omgeving)

Data Analytics Tool

Beschrijving

- ✓ Verzameling, integratie & transformatie van gestructureerde gegevens uit verschillende databronnen
- ✓ Geavanceerde statistische & AI-gestuurde analyses
- ✓ Automatische patroonherkenning & detectie van afwijkingen



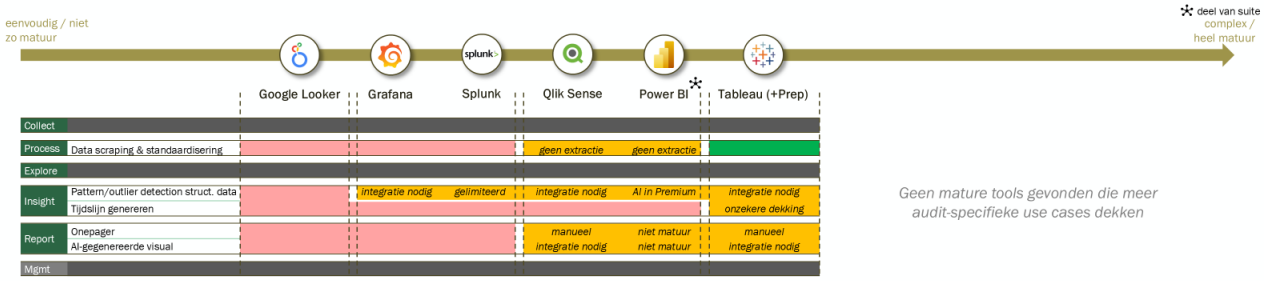
Voornaamste take-aways

- ➢ Er zijn kleine dedicated tools voor een gerichte functionaliteit (e.g. data-extractie & transformatie, netwerkanalyse, …)
- ➢ Daarnaast zijn er grotere platformen met een meer end-to-end ervaring, gebruiksvriendelijkheid/technische voorkennis van belang bij keuze
  - ➢ Er zijn zowel no/low-code tools (DataSnipper, Alteryx, IBM SPSS, Dataiku) als tools waarbij specifieke AI/ML/coding-voorkennis vereist is (Apache NiFi, Talend Data Fabric, RapidMinder, DataBricks, Anaconda)
- ➢ AI-visualisatie is zwakke plek van alle tools, al kan Anaconda veel via custom development

Data Visualization & Dashboarding Tool

Beschrijving

- ✓ Overzichtelijke visualisatie van data in grafieken/diagrammen, om inzichten te verwerven of voor gebruik tijdens de rapportering
- ✓ Dynamische & interactieve dashboards voor real-time inzichten



Geen mature tools gevonden die meer audit-specifieke use cases dekken

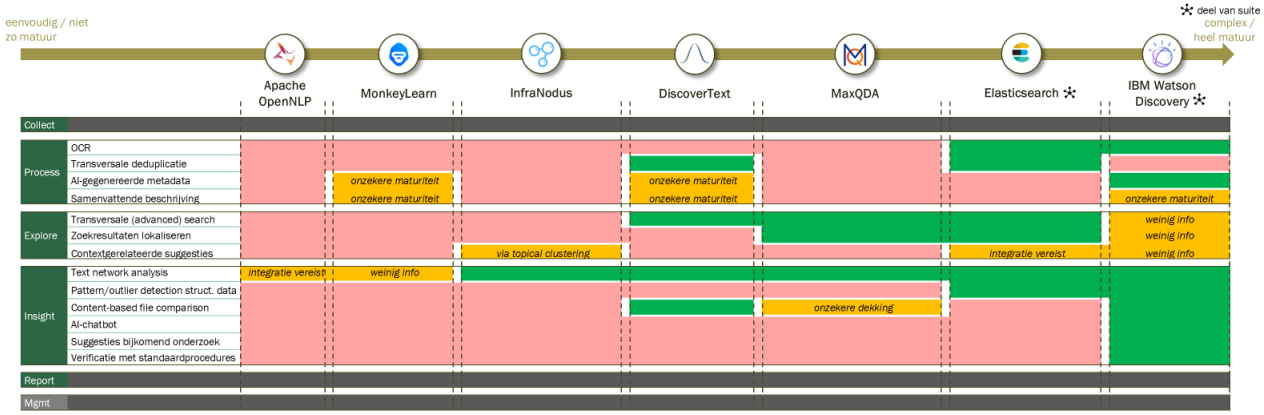
Voornaamste take-aways

- ➢ Alle tools kunnen met manuele input grafieken genereren voor rapportering, verschil zit vooral in flexibiliteit voor snelle interne visualisaties
- ➢ Geavanceerde/audit-specifieke visualisaties worden door geen enkele tool volledig ondersteund
- ➢ Sommige tools (Grafana, Splunk) focussen op dashboarding, andere (Power BI, Tableau) kunnen daarnaast ook gebruikt worden voor het genereren van stand-alone visuals

Text Mining/Analysis Tool

Beschrijving

- ✓ Extractie van inzichten/verbanden uit (grote hoeveelheden) tekst met behulp van sentimentanalyse, topic modeling, …
- ✓ Identificatie van trends, relaties & structuren in ongetructureerde (tekst)data



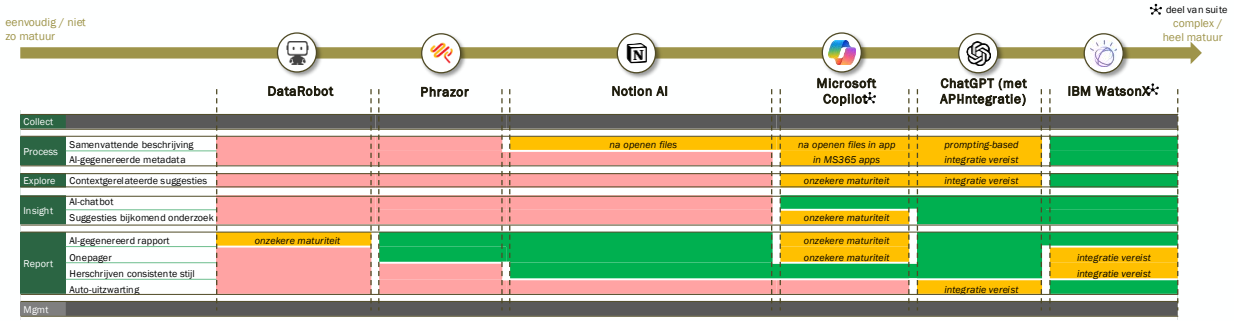
Voornaamste take-aways

- ➢ OCR slechts door beperkt aantal tools (< 1/3) ondersteund
- ➢ Meest geavanceerde tools (Elasticsearch, IBM Watson Discovery) bieden AI-functionaliteiten, maar vereisen specifieke voorkennis of maken deel uit van een suite/totaalpakket

Generative AI Tool

Beschrijving

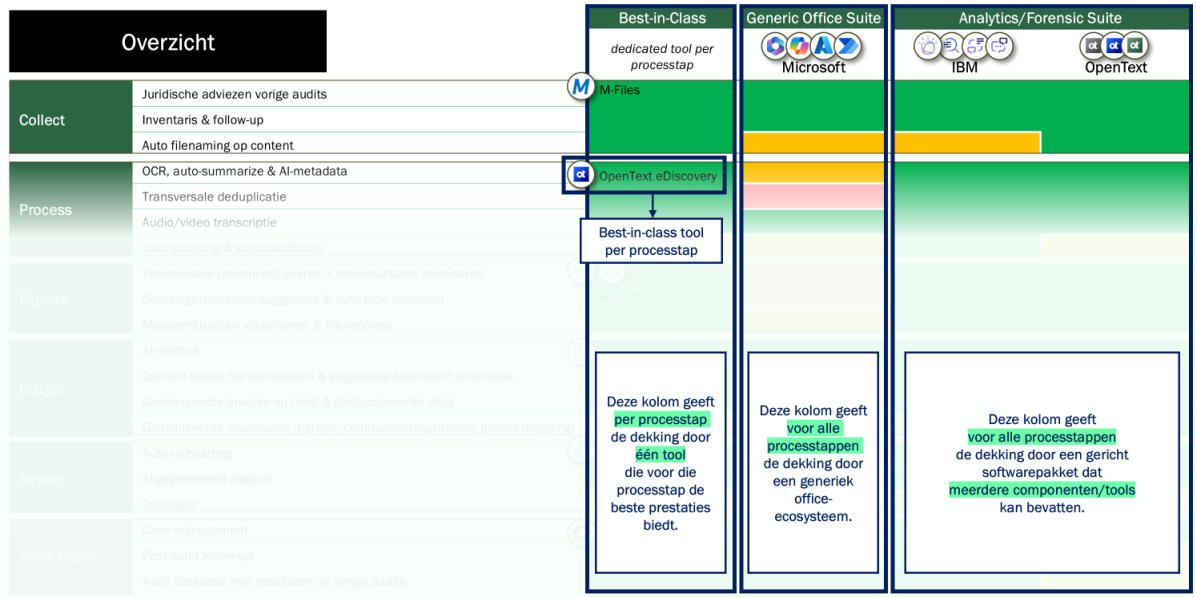
- ✓ Versneld genereren van samenvattingen, rapporteringen, …
- ✓ Ondersteuning voor inzichtsverwerving & analytische denkprocessen
- ✓ Doorzoeken van beschikbare data via interacties met een virtuele assistent of chatbot

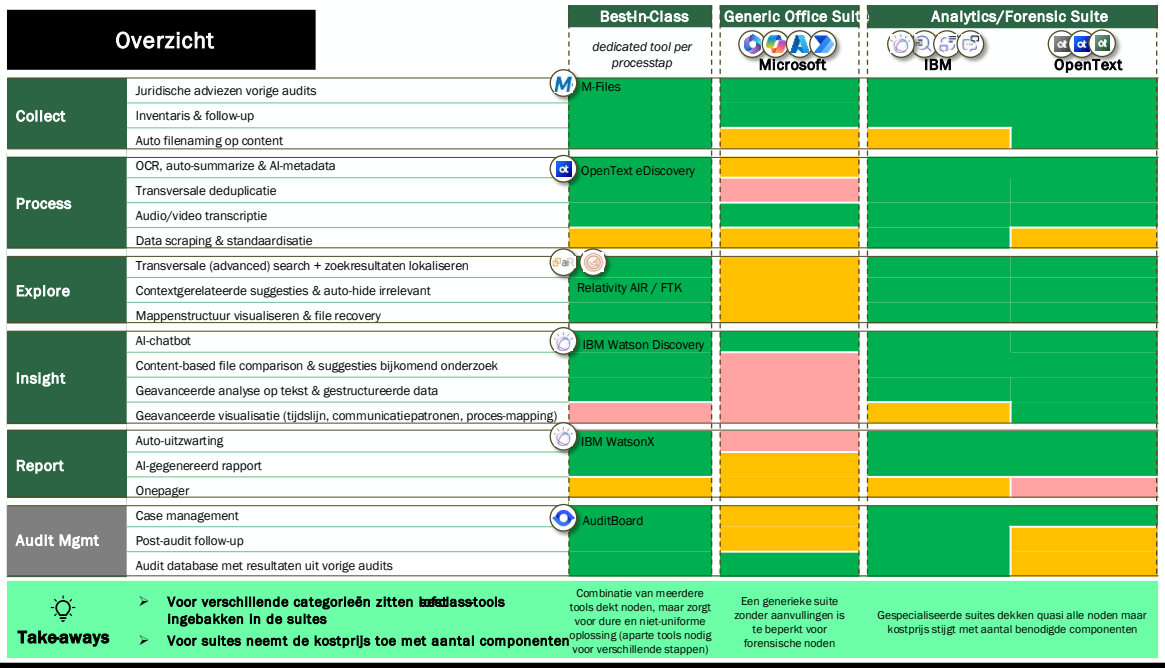


Voornaamste takeaways

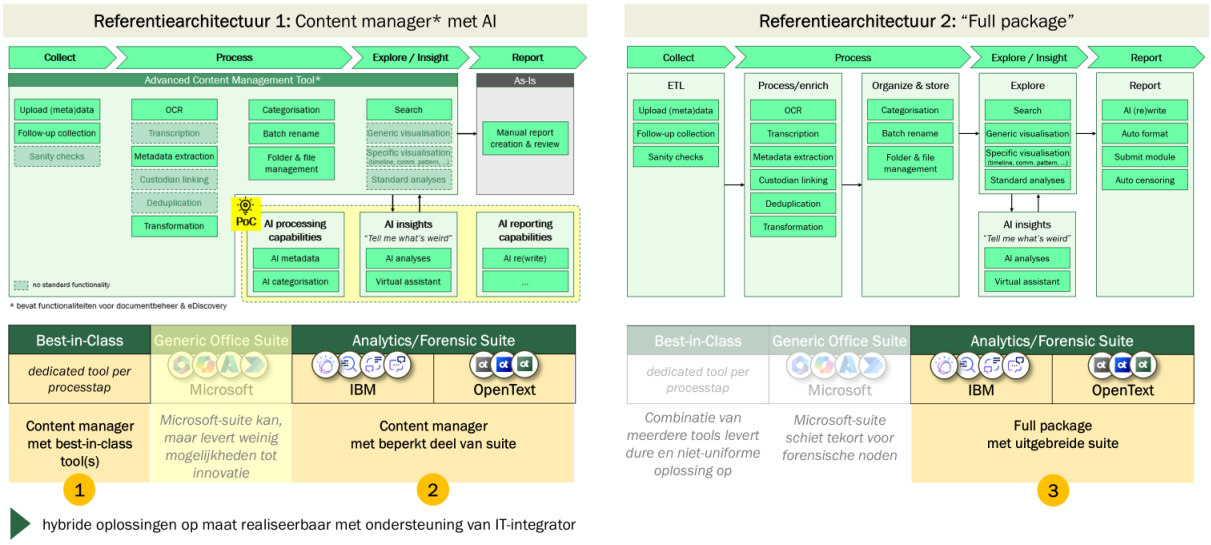
- ➢ Voor niet-geïntegreerde oplossingen hangt men af van de roadmap van de tool
- ➢ Geïntegreerde tools vereisen meer voorkennis, custom development & dedicated training (bv. rond prompt engineering)
- ➢ Confidentaliteit is vraagteken bij alle tools, wellicht zijn geïntegreerde toepassingen veiligste keuze

Overzicht





Combinatie referentiearchitecturen met resultaten marktanalyse levert 3 mogelijke keuzes op



Annex B: Deelnemers marktconsultatie

Volgende bedrijven schreven zich in voor de plenaire sessie van de marktconsultatie (op 24 juni 2025):

- ACA Group
- Auditstage
- **Avendar Technologies**
- B12 Consulting
- **Cegeka**
- **Cloudway**
- Codit
- **Craftzing**
- **Cronos Group**
- Delaware
- **Deloitte**
- Devoteam
- Dillen Technologies
- **DXC Technology**
- **element61**
- **EY**
- **Fujitsu**
- **IBM**
- **Inetum**
- **Intelli-Select**
- **KPMG**
- **ML6**
- **Möbius Business Redesign**
- **Moonland**
- Netcompany
- **Oor Services**
- **PwC**
- **Randstad Digital**
- **Robonext**
- Rollo.AI
- Sopra Steria
- **Spikes**
- **Valantic**
- **Youston**

De bedrijven waarmee nadien één-op-één-sessies werden gehouden zijn weergegeven in vet.