

Richtlijn Deduplicatie bij MedMij gegevensdienst Documenten

Versie: 0.1

Datum: 19-11-2024

Opgesteld door: Nictiz, productteam PGO

Inhoud en doel

In deze richtlijn wordt uitgelegd wat deduplicatie is en hoe dit kan worden toegepast bij de gegevensdienst Documenten. Deze richtlijn is bedoeld om bronsystemen te helpen dubbele documenten te voorkomen wanneer dezelfde documenten worden uitgewisseld. Daarnaast geeft de richtlijn ook uitleg over hoe PGO-leveranciers gedupliceerde documenten kunnen herkennen en verwijderen.

Ontwikkeling en Versie-beheer

Deze richtlijn is in samenwerking met MedMij ontwikkeld ten behoeve van de 3.0 publicatie van de gegevensdienst Documenten en is afgeleid van de Deduplicatierichtlijn ontwikkeld voor de 2.0.0-beta. 1 publicatie van de Informatiestandaard Vaccinatie-Immunisatie. Hoewel er in de standaard reeds handvatten worden aangeboden voor het voorkomen van dubbele documenten, zijn er nog geen (domein overkoepelende) richtlijnen vastgelegd voor het delen van Documenten en het uniek identificeren van gegevensobjecten. Dit document dient daarom als *tijdelijke* richtlijn specifiek voor de huidige Documenten publicatie. Zodra er Nictiz breed richtlijnen beschikbaar zijn opgesteld n.a.v. de projecten Documentuitwisseling en Objectidentificatie, zal deze specifieke Documenten Deduplicatierichtlijn hierop worden aangepast.

Deduplicatie

Deduplicatie (of ontdebellen) betekent het één keer tonen of overnemen van gedupliceerde gegevens.

Bij sommige informatietypen is het mogelijk om aan uniek identificerende gegevens te zien dat dit “hetzelfde” gegeven betreft. Deze kunnen ontdebeld worden, waarbij duplicaatinformatie niet getoond wordt.

Voorbeeld: Een cliënt gaat naar de huisarts met ernstige pijn aan de onderbuik. De huisarts verwijst de cliënt door naar het ziekenhuis voor een blinde darm verwijdering. Na de operatie wordt een ontslagbrief opgesteld door de specialist en deze wordt naar de huisarts gestuurd. Indien zowel het desbetreffende ziekenhuis als de huisarts als bronsysteem zijn aangesloten op de gegevensdienst Documenten, wordt deze ontslagbrief dubbel beschikbaar gesteld aan de PGO wanneer de cliënt al diens bestanden opvraagt.

Idealiter stellen partijen alleen hun *eigen* documenten beschikbaar aan de PGO, maar door gebrek aan harde richtlijnen wordt dit in de praktijk niet vaak nageleefd. Deduplicatie is daarom afhankelijk van de mogelijkheid tot automatische duplicaatdetectie: het vinden van duplicaatgegevens op basis van identificerende informatie. Indien deze niet of onvoldoende beschikbaar is, is het ook mogelijk handmatige deduplicatie toe te passen.

Toepassing in MedMij gegevensdienst Documenten

Automatische duplicaatdetectie is mogelijk op basis van het element Documentuitwisseling.Identificatie ([docs-dataelement-76](#)) / DocumentReference.masterIdentifier. Dit is de unieke documentidentificatie die wordt toegekend door de bron van het document. Deze identificatie geldt specifiek voor *deze versie* van het document.

Hoewel de gegevensdienst hiermee automatische duplicaatdetectie *mogelijk maakt*, betekent dit in de huidige praktijk niet dat er geen dubbele documenten kunnen voorkomen. Hieronder volgen twee scenario's waaruit dit blijkt:

Scenario A: Identificatie binnen één zorgsysteem

In sommige gevallen wordt de masterIdentifier pas aangemaakt bij het beschikbaarstellen aan de PGO. In dat geval kan het gebeuren dat als een patiënt twee keer hetzelfde document bij het ziekenhuis opvraagt, hetzelfde document twee keer beschikbaar wordt gesteld, elk met een andere masterIdentifier. Dit komt doordat het EPD mogelijk niet herkent dat het document al eerder is gedeeld met de PGO.

Scenario B - Uitwisseling tussen verschillende zorgaanbieders

Wanneer een zorgaanbieder (bijv. een huisarts) een PDF/A-document naar een ander zorgsysteem stuurt (bijv. een ziekenhuis), kan het gebeuren dat elk systeem zijn eigen masteridentifier toekent aan het document. Bij de uitwisseling met een PGO is het niet duidelijk genoeg vastgelegd dat alleen zelf gegenereerde documenten mogen worden uitgewisseld. Hierdoor kan de situatie ontstaan waarin de cliënt zowel bij het ziekenhuis als bij de huisarts een document opvraagt en beide hetzelfde document sturen, maar elk met een andere masteridentifier. Door de verschillende identifiers is automatische deduplicatie niet meer mogelijk en kunnen geduplicateerde documenten ontstaan.

Handmatige duplicaatdetectie: gebruik metadata

Handmatige duplicaatdetectie kan met ondersteuning van de PGO-dienst plaatsvinden. Indien er documenten worden getoond die overeenkomstige meta-informatie bevatten, kan de PGO-gebruiker via een attentie (bijvoorbeeld via een uitroepteken, specifieke kleur of andere vorm) worden gewaarschuwd voor een *mogelijke* dubbele weergave van hetzelfde document. De PGO-leverancier kan vervolgens de PGO-gebruiker de mogelijkheid bieden om één van dezelfde documenten in het overzicht te verbergen.

De trigger voor deze attentie is niet officieel vastgelegd maar zou *kunnen* bestaan uit de combinatie van (verplichte) metadata-elementen zoals een identieke IndexDatum (1..1R) met een identiek Type (1..1R) en identieke Titel (0..1R).

Scenario A - Identificatie binnen één zorgsysteem

Omdat de duplicaten uit dezelfde bron komen en daardoor identieke inhoud en meta-informatie hebben, kunnen metadata elementen (bijvoorbeeld Titel, IndexDatum) worden gebruikt om te bepalen dat het om hetzelfde document gaat.

Scenario B - Uitwisseling tussen verschillende zorgaanbieders

In deze situatie is het lastiger om te bepalen of er voldoende andere metadata beschikbaar is om te identificeren dat het om hetzelfde document gaat aangezien elk systeem eigen/andere metadata kan invullen.

Voorwaarden voor betrouwbare deduplicatie

Zowel voor automatische als handmatige duplicaatdetectie is het cruciaal dat alle betrokken leveranciers en zorgaanbieders de masterIdentifier op een consistente manier gebruiken en de verplichte metadata-elementen volledig en correct invullen.

Hoewel zowel de globale MedMij Implementation Guide (IG) als de PDF/A IG het gebruik hiervan aanbevelen, zijn er geen verplichte implementatierichtlijnen. Dit leidt er in de praktijk vaak toe dat essentiële meta-informatie ontbreekt, wat het moeilijk maakt om duplicaten automatisch of handmatig te identificeren.

Om betrouwbare duplicaatdetectie mogelijk te maken, zijn zorgbrede afspraken nodig over het gebruik en de uitwisseling van unieke identifiers bij documenten.

Toekomstige ontwikkelingen

Een generiek vraagstuk wat momenteel Nictiz breed onderzocht wordt, is hoe gegevensobjecten uniek geïdentificeerd kunnen worden en welke Nictiz brede richtlijnen daarbij horen. Daarnaast wordt ook onderzocht welke domeinoverstijgende afspraken nodig zijn voor het optimaliseren van documentuitwisseling, zowel tussen zorgverleners als tussen zorgverleners en patiënt. Zodra Nictiz brede richtlijnen en afspraken zijn vastgesteld, zullen de huidige deduplicatierichtlijnen voor Documenten worden hierop worden aangepast.