

# 1 Post-FAIRificatie

De laatste fase is de post-FAIRificatie. Hierin volgt in stap 6 een evaluatie. In stap 7 en 8 worden er aanbevelingen gedaan voor het automatisch produceren en uniform publiceren van de Woo-dossiers.

## 1.1 Stap 6: Evalueren van de FAIRificatie

Om de kwaliteit van de Woo-dossiers na FAIRificatie te kunnen evalueren zal er allereerst een voorbeelddossier worden omgezet. Als voorbeeld zal een reeds gepubliceerd dossier van de provincie Noord-Holland worden gebruikt. In het licht van de vervanging van de Wob door de Woo is de provincie een pilot gestart waarbij er zoveel mogelijk informatie actief openbaar wordt gemaakt met betrekking tot de plaatsing van een zonneweide aan de Jaagweg. In de pilot wordt er gezocht naar een manier om informatie binnen één project openbaar te maken. In de huidige situatie worden de dossiers echter nog op dezelfde wijze gepubliceerd als bij de Rijksoverheid, vergelijkbaar met de bevindingen van de analyse van de dossiers in stap 1. Een FAIRificatie zal worden uitgevoerd op een dossier uit het project met daarin de maanden september tot en met december in 2021. De publicatie van dit dossier is te vinden in Figuur 1, waarin simpelweg de pdf documenten worden gepubliceerd. In Codevoorbeeld 1 staat een fragment van een JSON document gemaakt van het bovengenoemde dossier, op basis van het in stap 3, 4 en 5 opgestelde schema met metadata en restricties. De waarden zijn grotendeels gevonden in de documenten uit het dossier, zoals een titel, besluitdatum en een beoordeling. Andere waarden, zoals de identifiers, een thema, de bodytext en aantallen woorden, karakters en pagina's zijn zelf bedacht of automatisch geëxtraheerd uit de documenten met Python.

**Figuur 1:** Dossier uit de pilot actieve openbaarmaking van de provincie Zuid-Holland.

### Zonneweide Jaagweg (sept, okt, nov, dec 2021)

[Besluit actieve openbaarmaking Zonneweide Jaagweg sept okt nov dec 2021](#)  
(23 december 2021, pdf, 159kB)

[Inventarisatielijst sept okt nov dec 2021 dossier zonneweide Jaagweg](#)  
(23 december 2021, pdf, 68kB)

[1. Memo BO verslag gesprek 20210324](#)  
(23 december 2021, pdf, 150kB)

[2. Memo BO verslag gesprek 20210421](#)  
(23 december 2021, pdf, 162kB)

[3. Memo BO verslag gesprek 20210520](#)  
(23 december 2021, pdf, 160kB)

[4. Memo BO verslag gesprek 20210907](#)  
(23 december 2021, pdf, 178kB)

[5. Agenda BO 18112021 gedeputeerde Stigter en Koggenland zonneweide Jaagweg](#)  
(23 december 2021, pdf, 157kB)

**Codevoorbeeld 1:** Fragment van het gecreëerde JSON document van het Zonneweide Jaagweg dossier.

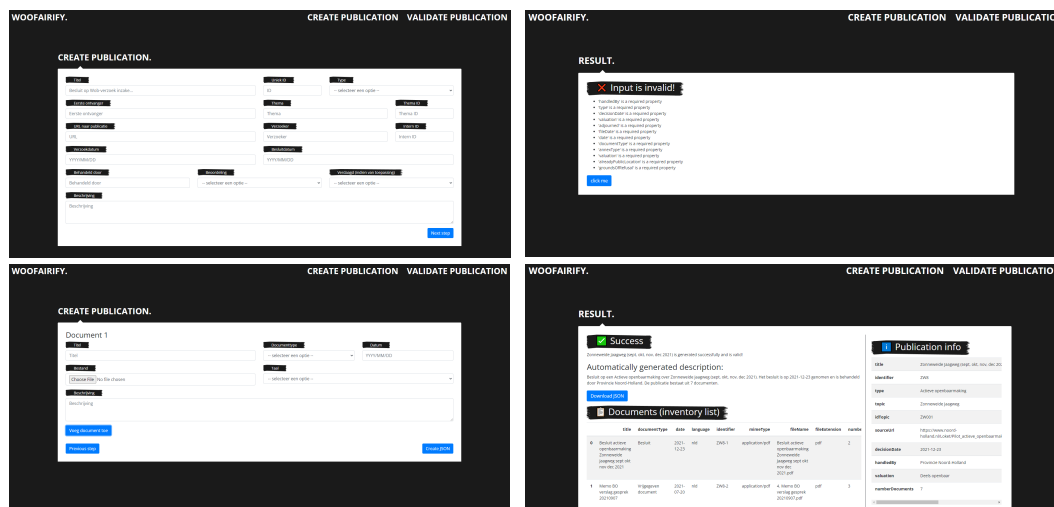
```
{
  "title": "Zonneweide Jaagweg (sept, okt, nov, dec 2021)",
  "identifier": "Woo.ZW8",
  "type": "Actieve openbaarmaking",
  "topic": "Zonneweide Jaagweg",
  "idTopic": "ZW001",
  "decisionDate": "2021-12-23",
  "handledBy": "Provincie Noord-Holland",
  "valuation": "Deels openbaar",
  "numberDocuments": 7,
  "documents": [
    {
      "title": "Memo BO verslag gesprek 20210520",
      "documentType": "Vrijgegeven document",
      "date": "2021-07-19",
      "language": "nld",
      "recipient": "Gedeputeerde Stigter, wethouder Van Dolder",
      "annexType": "Memo",
      "valuation": "Deels openbaar",
      "groundsOfRefusal": [
        "10.2.e"
      ],
      "identifier": "Woo.ZW8.7",
      "mimeType": "application/pdf",
      "fileName": "3. Memo BO verslag gesprek 20210520.pdf",
      "fileExtension": "pdf",
      "bodyText": "Betreft: Verslag bestuurlijk overleg met
                    wethouder Van Dolder
                    ...
                    Communicatie
                    Geen onderwerpen om te bespreken."
      "numberPages": 2,
      "numberCharacters": 3031,
      "numberWords": 449,
      "numberTextPages": 2
    },
    ...
  ]
}
```

Op basis van dit omgezette voorbeeld kunnen de opbrengsten per FAIR Data principe in kaart worden gebracht. In tabel 1 is deze evaluatie te zien. Per FAIR principe is de situatie vóór, zoals gevonden in de eerste stap, en ná FAIRificatie geschetst om aan te geven hoe de FAIRificatie dit dossier heeft verbeterd op dat principe. Op elk principe heeft er een postieve verandering plaatsgevonden.

**Tabel 1:** Evaluatie van het FAIRificatie-resultaat.

FAIR Principe	Voor FAIRificatie	Na FAIRificatie
Vindbaarheid	Dossiers en documenten missen relevante metadata en/of deze zijn slecht vindbaar. Inventarislijsten zitten soms ‘verstopt’ en de pdf documenten zijn niet goed te doorzoeken.	Dossiers staan opgeslagen in een goed doorzoekbaar formaat (JSON). De toevoeging van identifiers en het toekennen van een thema dragen bij aan het verhogen van de vindbaarheid van dit dossier. Het extraheren van de bodyText maakt het mogelijk om documenten te kunnen doorzoeken met een zoekmachine.
Toegankelijkheid	Alle documenten worden in een pdf formaat verstrekt, waarvan bijna 25% niet toegankelijk en leesbaar is voor een computer. Het is lastig om automatisch bruikbare informatie te vinden in de pdf documenten.	Het dossier is mens- en machine-leesbaar door het JSON formaat. Het formaat maakt het mogelijk waarden eenvoudig te extraheren en verder te verwerken met een programmeertaal of zoekmachine-tool.
Uitwisselbaarheid	Ieder bestuursorgaan heeft een eigen wijze van labelen en publiceren. Dit resulteert in ambiguïteit voor zowel attribuutnamen als waarden.	Door te voldoen aan strenge validatie-eisen voldoet dit dossier, net als andere dossiers, aan een minimale standaard. Metadata worden op een uniforme wijze toegekend.
Herbruikbaarheid	De dossiers missen vindbare en toegankelijke metadata, wat de herbruikbaarheid voor onderzoek en monitoring niet ten goede komt. Ook voor de bestuursorganen zelf zorgt het gebrek aan metadata waarschijnlijk voor slechte archivering en herbruikbaarheid.	Door de identificatie van metadata vanuit het perspectief van zowel handelaars als gebruikers van de dossiers is het uniek te identificeren dossier met documenten herbruikbaar voor zowel toekomstige Wooverzoeken als journalistiek of wetenschappelijk onderzoek.

**Figuur 2:** Screenshots van de ontwikkelde Woo-tool WooFAIRify.

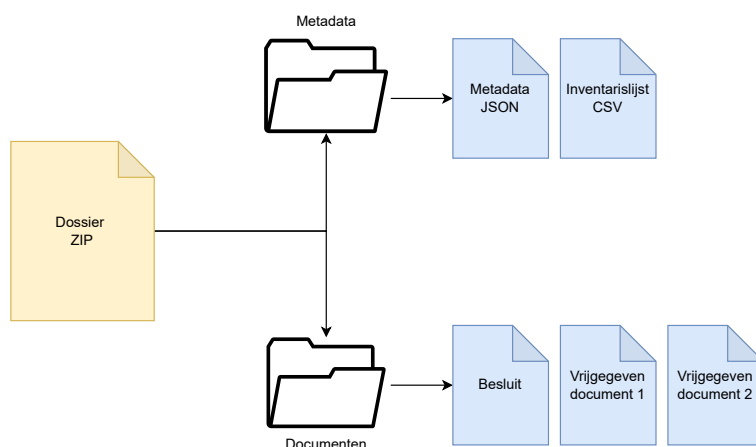


## 1.2 Stap 7: Het automatiseren van FAIR produceren

Met het behalen van de eerste twee doelen, het vinden van een leesbare kennisrepresentatie en het identificeren van relevante metadata, blijft het laatste doel nog over: het ontdekken van een automatische en uniforme wijze van labelen en publiceren. In deze stap zal in worden gegaan op het automatiseren van de productie van Woo-dossiers op een FAIR manier. Het produceren van dossiers is een routineklus en moet niet te moeilijk zijn en te veel tijd kosten. Voor dit doeleinde is er een tool ontwikkeld, 'WooFAIRify' (?). Het doel van WooFAIRify is het bieden van een eenvoudige interface voor het invullen van relevante informatie waarna de gebruikersinput wordt gevalideerd en omgezet naar het wensbare formaat volgens de FAIRificatie, in dit geval JSON.

In Figuur 2 zijn de vier belangrijkste schermen weergegeven. Een behandelaar vult eerst in het scherm linksboven de gegevens over een dossier in, zoals een titel en een datum. Daarna kan de behandelaar documenten toevoegen (zie linksonder) met ook daarbij de metadata. Vervolgens worden de ingevulde gegevens gevalideerd. Als dat mislukt krijgt de behandelaar feedback op de ingevulde gegevens. In het geval van het scherm rechtsboven is bijvoorbeeld de behandelaar (handledBy) niet verstrekt. Als de ingevulde gegevens helemaal voldoen aan de opgestelde regels en beperkingen uit stap 5, verschijnt het resultaat in beeld. Op dit scherm kunnen een JSON van het gehele dossier en een CSV van de documenten als een inventarislijst worden gedownload. De tool maakt het mogelijk om zonder technische kennis van de attributen en het datamodel tóch op een uniforme manier dossiers te produceren.

**Figuur 3:** Voorbeeld van de publicatiestrategie.



### 1.3 Stap 8: Identificeren van een publicatiestrategie

De laatste stap van het FAIRificatie proces geeft een handreiking voor het op een uniforme wijze publiceren van de automatisch gemaakte dossiers. Het is wenselijk om een Woo-dossier te voorzien van alle documenten die zijn vrijgegeven in pdf formaat. Daarnaast is het resultaat van de FAIRificatie een semigestructureerd document met daarin vindbare, toegankelijke, uitwisselbare en herbruikbare metadata. Eventueel kan er ook nog een inventarislijst in tabelvorm worden toegevoegd aan het dossier. Om al deze documenten te bundelen in een herkenbare vorm is het noodzakelijk om een publicatiestrategie op te stellen. Figuur 3 doet een voorzet voor een publicatiestrategie, waarin het gehele dossier als een ZIP bestand wordt verstrekt. Deze ZIP bevat vervolgens twee mappen: één map met daarin de metadata, het resultaat van de FAIRificatie, en een andere map met daarin de originele pdf documenten als reactie op een informatieverzoek of een actieve openbaarmaking.

#### Voorbeeld Zonneweide Jaagweg

Het in Figuur 1 gebruikte dossier kan met WooFAIRify simpel worden omgezet naar het gewenste formaat zoals in Codevoorbeeld 1. Figuur 4 laat zien hoe de tool gebruikt kan worden om alle relevante en beschikbare gegevens over het dossier en bijbehorende documenten in te vullen. De tool bevat ook een ingebouwd voorbeeld op basis van hetzelfde dossier. Het huidige dossier mist unieke identifiers, dus deze zijn toegevoegd als dummy-waarden. Het thema (topic) kan worden ingevuld om aan het doel om informatie binnen één project openbaar te maken te voldoen. Voor de documenten kunnen er bestanden worden toegevoegd, waarna de tool automatisch velden als de extensie, de letterlijke tekst, het aantal woorden, karakters en pagina's extraheert. Een fragment van het gemaakte JSON document is te zien in Codevoorbeeld 1.

**Figuur 4:** Gegevens over een dossier van Zonneweide Jaagweg ingevuld in WooFAIRify.

The screenshot shows the WooFAIRify web application interface. It features a form for entering dossier information. The form is organized into two main sections: 'Document 1' and 'Documents 2-6...'. Each section contains fields for 'Title', 'Document type', and 'Date'. Below these fields, there is a 'Choose File' button and a 'Description' field. The 'Document 1' section is filled with data for a dossier titled 'Zonneweide Jaagweg' (sept, okt, nov, dec 2021). The 'Documents 2-6...' section is also filled with data for a dossier titled 'Zonneweide Jaagweg' (sept, okt, nov, dec 2021). The form includes various dropdown menus and text input fields for detailed information.

Met de door de tool gemaakte JSON en CSV documenten kan er een voorbeeldpublicatie van het dossier worden gemaakt. In Figuur 5 is deze te zien. In twee simpele stappen - het invullen van de benodigde informatie in de tool en het creëren van een zip bestand met een overzichtelijke mappenstructuur - is er een volledig dossier gecreëerd, klaar voor publicatie.

**Figuur 5:** Boomstructuur van de mappen en documenten behorende bij het Zonneweide Jaagweg dossier.

```

ZW8.zip
├── Metadata
│   ├── ZW8-data.json
│   └── ZW8-inventaris.csv
└── Documenten
    ├── Besluit actieve openbaarmaking Zonneweide Jaagweg sept okt nov dec 2021.pdf
    ├── Inventarisatielijst sept okt nov dec 2021 dossier zonneweide Jaagweg.pdf
    ├── 1. Memo B0 verslag gesprek 20210324.pdf
    ├── 2. Memo B0 verslag gesprek 20210421.pdf
    ├── 3. Memo B0 verslag gesprek 20210520.pdf
    ├── 4. Memo B0 verslag gesprek 20210907.pdf
    └── 5. Agenda B0 18112021 gedeputeerde Stigter en Koggenland zonneweide Jaagweg.pdf
    
```