

## Vormgeving van de Nederlandse Data Agenda: relevantie voor eRecht

### Data science for the good

Tijdens de Haagse [Hackaton for peace, justice and security](#) - in kort "Hackaton for good" - legde een aantal deelnemende partijen vragen voor aan overwegend enthousiaste jonge geeks. Een van de [challenges](#) kwam van het Internationaal Strafhof (ICC).

Het probleem dat het Internationaal Strafhof neerlegde was het volgende. De procureur van het hof analyseert een groot aantal documenten in geschreven vorm van velerlei oorsprong: nieuwsartikelen, NGO rapporten, getuigenverklaringen, etc. Deze teksten bevatten informatie over plaatsen en personen. De vraag van de datascientist van het Hof was, hoe kunnen de relaties tussen de in de documenten genoemde personen en plaatsen automatisch worden vastgesteld. Dit is relevant, aangezien het Hof nu gebruik maakt van analisten die manueel dit werk uitvoeren. De verwachting van de *challenge* was dat deze taak goedkoper, sneller en beter uitgevoerd kan worden door de computer. Maar, was vervolgens de vraag, hoe pakken we dat dan aan?

Enkele dagen volgend op de Hackaton vond – eveneens in Den Haag, het congres plaats met de titel: [Data Science voor maatschappelijke uitdagingen](#). De conferentie was georganiseerd door het [Leiden Centre of Data Science](#) en [Science Works](#) in het kader van de inspanning van de Nederlandse overheid om een [Nationale Data Agenda](#) te formuleren. Om te assisteren in de vormgeving waren verschillende – veelal multidisciplinaire, datacentra van de Nederlandse universiteiten aanwezig om met de overheid te sparren over wat er zeker niet in de data-agenda mag ontbreken.

### Data-infrastructuur, ook voor het recht

De overheid moet helpen om een goede infrastructuur neer te leggen waarin data gebruikt kan worden om processen over te nemen waarin machines veel beter zijn dan mensen, zo werd tijdens het congres eensluidend gesteld. Ons land doet dit door een belangrijke rol te spelen in het initiatief dat de architectuur vormgeeft voor data-gebruik door machines. Deze architectuur draagt het acronym FAIR, wat staat voor Findable, Accessible (*under well-defined conditions*), Interoperable en Reusable. Deze architectuur heeft allereerst tot doel om data leesbaar te maken voor machines, omdat alleen computers via algoritmes in staat zijn de giga of terabytes aan data door te nemen en daaruit verbanden te destileren.

Dit is heel relevant – ook voor het rechtendomein. De tekst-documenten, zoals het Internationaal Strafhof stelde in haar *challenge*, zijn bij uitstek niet geschikt voor *machine-readable* analyse. Daarin moeten allereerst de kern-concepten bepaald en gevonden worden, waarna pas de relaties tussen deze concepten kunnen worden onderzocht. De relaties die vervolgens gelegd worden, geven belangrijke aanwijzingen voor hypotheses, die vervolgens verder onderzocht kunnen worden door de juristen.

### Internet der data

Het FAIR protocol is de opstap naar het internet van data omdat het in een data-management voorziet op basis van overeengekomen principes die er voor zorgen dat de data door machines ingelezen kunnen worden. Dat maakt de data dus 'open'. Dat betekent overigens niet dat de data onder alle omstandigheden 'open' zijn. In feite wordt de data

gesegmenteerd en wordt via meta-data zichtbaar, waardoor verschillende onderdelen van de data beschikbaar worden, mits toestemming is verkregen. FAIR voorziet dus in een veilige architectuur dat de toestemming voor *data-entry* regelt middels toestemming door het *data-subject*. Daarmee heeft de overheid een belangrijke rol om via deze architectuur het commerciële gebruik van data te registreren, door gebruik te verbinden aan toestemmings-verplichting. Hiermee wordt het publieke belang gediend.

### **Het internet terug naar de *data subjects***

Niet alleen voor het toekomstige internet van data is de bescherming van data en *data subjects* van belang, maar ook voor het huidige internet. [Sir Nigel Shadbolt](#), de rechterhand van [Tim Berners-Lee](#) (de man achter het *world wide web*), presenteerde een initiatief [SOLID \(Social Linked Data\)](#) om tot een soortgelijke bescherming te komen van het internet. Het internet, zo stelde Sir Nigel, moet terug gegeven worden aan de *data subjects* die zich vervolgens in *trusts* kunnen verenigen om zo de toestemming voor het gebruik van hun data te bepalen. De hoop is dat zo het publiek meer vertrouwen krijgt in het gebruik van data voor het oplossen van maatschappelijke uitdagingen.

### **Data science voor het (open)recht**

De *Hackaton-for-Good* in Den Haag werd enthousiast afgesloten met 23 prachtige *elevator-pitches*. De informatici van het ICC lieten zich inspireren door de gedrevenheid waarmee de jonge *computer scientists* aan de slag gingen om een oplossing te geven voor een probleem dat, indien opgelost, handen vrij kan maken voor het 'echte' onderzoek van juristen naar de vervolging van internationale strafzaken.

Het congres in Den Haag eindigde eensgezind, met de aanbeveling dat de Nederlandse overheid moet investeren in een goede infrastructuur voor data science. Bovenal moet de overheid de architectuur omarmen van Nederlandse bodem: FAIR. Dit biedt immers een perfecte basis voor het oplossen van *nontrivial data-problems*, zoals de computer scientist het zou uitdrukken.

Deze *events* hebben richting gegeven aan de ontwikkeling van data science, een data-infrastructuur, een internet der data. Een ontwikkeling die ook van groot belang is voor het recht en waarbij OpenRecht een grote rol kan gaan spelen als leverancier van *FAIR legal data*.

### **[Mirjam van Reisen](#)**

Professor of Computing for society  
Universiteit Leiden