

# De Andere Helft. Expertmeeting #5: De grote data en registratie-expertmeeting

14 oktober 2022 – 10:00-13:00

*verslag door Bert-Jaap Koops*

## Overzicht

Samenvatting. Conclusies en suggesties .....	2
Wat kun je allemaal met je data doen? .....	2
Hoe begin je?.....	2
Waar moet je op letten? .....	2
Uitgebreid verslag. Inleiding.....	3
Bram Donders (Museum Boijmans Van Beuningen) en Mayken Jonkman (RKD) – Welkom en intro	3
Deel I: Wat kunnen we met data? Wat voor onderzoeksresultaten worden behaald? .....	3
Weixuan Li (Huygens Instituut/UvA) – Space and Strategy: Mapping 17th-century art market in Amsterdam .....	3
Harm Nijboer (UvA) en Judith Brouwer (Huygens Instituut) – Onderzoek in een ogenblik. Prentkunst in Amsterdam in de zeventiende eeuw .....	5
Algemene vragen en discussie .....	6
Deel II: Linked Open Data en Praktijk: Wat wordt er al gedaan en wat kunnen wij doen? .....	7
Reinier van 't Zelfde (RKD) – Door het LOD verbonden aan het RKD .....	7
Edwin Klijn (Netwerk Oorlogsbronnen) – Van archieven naar verhalen. 1 + 1 =3 .....	10
Daniel Kapitan (EAISI) – Knowledge Graphs in de praktijk.....	11
Algemene vragen en discussie .....	12
Afsluiting.....	13

## Samenvatting. Conclusies en suggesties

### ***Wat kun je allemaal met je data doen?***

- **Je kunt heel veel met data, mits je ze goed digitaliseert, opslaat volgens veelgebruikte standaarden, annoteert met behulp van geharmoniseerde woordenboeken, en ontsluit via FAIR-principes (vindbaar, toegankelijk, interoperabel en herbruikbaar).**
- Door data in een gestructureerde databank te plaatsen, kun je ze analyseren en onvermoede verbanden ontdekken. Dat kan nog beter als je data uit verschillende bronnen samenvoegt. Delen is dus belangrijk.
- Door data uit verschillende bronnen te koppelen, kun je nieuwe verhalen vertellen. Namen worden een persoon als je verschillende soorten gegevens uit uiteenlopende databronnen weet te combineren.
- Je kunt data uit databanken analyseren en de uitkomst visualiseren in grafieken of kaarten. Combinaties van data (bijvoorbeeld: waar woonden kunstenaars en waar woonden hun afnemers) kunnen verbanden inzichtelijk maken. Een plaatje of kaart zegt meer dan vele woorden.
- Gebruik handige functionaliteiten die programma's of applicaties hebben, zoals schaalbare kaarten waarop je kunt in- en uitzoomen.
- Denk ook aan andere manieren om je data te organiseren; met kennisgrafen kun je vaak meer dan met een klassieke (Excel-achtige) databank.

### ***Hoe begin je?***

- **Klein.** Maakt het niet te ingewikkeld, maar begin met een beperkt aantal eigenschappen of een overzichtelijke deelverzameling. Je kunt vervolgens geleidelijk uitbreiden.
- Begin gewoon. Al doende leert men.
- Denk wel goed na hoe je dingen gaat vastleggen, in de juiste context, op een standaardmanier. Wat en hoe je vastlegt, hangt ook af van wat je met de data wil gaan doen.
- Verdiep je in bestaande databanken; vaak bestaat al meer dan je denkt. Wikidata bevat zelf veel bruikbare data, maar heeft onderaan de pagina ook handige links naar meer databanken.
- Laat je niet afschrikken door allerlei technische afkortingen, programma's en formaten. Als kunt leren de weg in het Frans te vragen, kun je ook leren zoekvragen in SPARQL te formuleren. Het loont de investering je erin te verdiepen, en al doende leert men.
- Stel je data beschikbaar als **Linked Open Data**, door ze in breed gebruikte standaard-formaten op te slaan en beschikbaar te stellen voor anderen.
- Sluit aan op bestaande databanken (RKD Explore, Ecartico, Wikidata) om nieuwe data (of nieuwe verbanden) te ontsluiten, in plaats van zelf iets te ontwikkelen. Daar is immers al over nagedacht, en zo bereik je snel meer mensen.

### ***Waar moet je op letten?***

- **Gebruik standaarden.**
- Gebruik open technische standaarden (bijvoorbeeld voor Linked Open Data). Gebruik veelgebruikte terminologische standaarden en ontologieën.
- **Zie data los van de applicatie waarin ze toevallig ooit zijn gegoten.** Het maakt niet zoveel uit welk beheersysteem (Axiell, TMS) je gebruikt; belangrijker is dat je data op een gestandaardiseerde manier vastlegt, zodat ze koppelbaar zijn.
- Gebruik zoveel mogelijk *persistent identifiers* (vaste URL's of webadressen), maar realiseer je dat ook die onderhoud vergen.
- Maak werk van betrouwbare data. Betrouwbare data zijn data die geschoond zijn, gegrond in gedeelde begrippen, gecontextualiseerd en toegankelijk in een gestandaardiseerd formaat, en die worden beheerd en geactualiseerd.

- Houd het onderscheid tussen brongegeven en annotatie in de gaten: historici maken immers een scherp onderscheid tussen brongegeven en interpretatie.
- Als blijkt dat je een naam of kunstwerk dubbel hebt ingevoerd, laat dan de ene bestaan (anders werken verwijzingen ernaar niet meer) maar link deze automatisch door naar de andere.
- Werk samen met andere instellingen. Beschouw je eigen data niet als zelfstandige data van jezelf, maar als onderdeel van een breed kennisnetwerk. Deel met anderen en maak gebruik van door anderen gedeelde data.
- Vraag tips aan collega's (ook bij andere instellingen) die al met bepaalde bijtjes hebben gehakt. Help elkaar verder.
- Voorkom dubbel werk: doordachte koppeling van data is beter dan zelf data opnieuw invoeren.
- Staar je niet blind op technische toepassingen en de mogelijkheden van geautomatiseerde data-analyse. Expertise blijft belangrijk om de juiste vragen te stellen en resultaten te interpreteren. Veel kan worden geautomatiseerd, maar handwerk blijft altijd nodig.
- Begrippen zijn belangrijk. Begrip van begrippen is nog belangrijker.
- Data over vrouwen kun je niet los zien van data over mannen. Ze behoren gezamenlijk tot netwerken.

## **Uitgebreid verslag. Inleiding**

### ***Bram Donders (Museum Boijmans Van Beuningen) en Mayken Jonkman (RKD) – Welkom en intro***

Voor mensen die niet aanwezig konden zijn en graag willen terugkijken, wordt de bijeenkomst opgenomen; sprekers zijn hiermee akkoord. De opname is tijdelijk te bekijken op [link is niet meer beschikbaar i.v.m. privacy].

Aanleiding voor deze bijeenkomst was de vaak opkomende vraag: hoe kunnen we onze data delen? Het kan innovatiever dan een gezamenlijke Google Docs. Een oproep op LinkedIn om initiatieven door te geven leidde tot diverse gesprekken en uiteindelijk tot deze expertmeeting. Informatie-uitwisseling leeft zeer, ook binnen de gemeenschap van professionals die zich bezig houden met vrouwen in de kunst tussen 1780-1980. Waar zien we potentie, hoe kunnen we dit aanpakken?

## **Deel I: Wat kunnen we met data? Wat voor onderzoeksresultaten worden behaald?**

### ***Weixuan Li (Huygens Instituut/UvA) – Space and Strategy: Mapping 17th-century art market in Amsterdam***

*Weixuan Li is finishing her PhD at the Huygens Institute and the University of Amsterdam, as part of the [Virtual Interiors as Interfaces for Big Historical Data Research](#) project. Her research aims at contextualizing artists' lives and working environments in 17<sup>th</sup>-century Amsterdam through multi-layered deep mapping. Li specializes in digital art history and applies computational methods to understand the development of early modern art and the art market.*

This presentation illustrates how maps can be used to tell stories about the past. This research, which focuses on artists in the 17<sup>th</sup> century, is not particularly about women, but can well be applied to it. Li uses the semantic web and 2D mapping to create a socio-spatial perspective on the art market in an immersive user experience. Using various sources, including the visual database of RKD, the descriptions of social networks in Ecartico, Amsterdam city archives, and cadastral maps of HISGIS, it is possible to 'deep-map' the complex 17<sup>th</sup>-century Amsterdam art market. The maps are scalable



and allow following the temporal pattern of painters' careers over a period of over a hundred years.

Li sketches the development of the art market in 17<sup>th</sup>-century Amsterdam and illustrates this with maps of artists, patrons and other actors in the art market. Reputation and visibility were key; hence artists often established themselves close to where patrons lived. Immigrants from the Southern Netherlands instigated art trade. They were very well-connected, both physically and socially, in different clusters, as visualised in a network map. Artistic centres expanded, e.g., around Jodenbreestraat and Sint Antoniebreestraat. For outsiders, cultivation of relationships was more difficult; they resorted to various integration strategies.

Visualising the spread of sitters of portrait painters shows similarity to the spread of wealth in the city. With population growth and house walls filling up with paintings, the market expanded and painters specialised in their own style. Lower-tier painters resided close to market squares, so they could sell without strong ties to patrons. Some artists moved upwards, such as Karel Dujardin, visible in their residential moves across the city. In 1650-1670, when the population reached its peak, there were three main artistic centres; with subsequent declining population, artists spread out.

The presentation thus shows how the cycle of developments in the art market can be visualised in maps. Elaboration and further examples can be found in Li's upcoming dissertation.

## Questions

Q: Which sources did you use and how did you combine them?

A: I used data from Ecartico about painters' lives and social networks. Other sources, e.g., choices artists made in their development, came from art historical literature. So data and art historical information have to be combined.

Q: Did you create your own database?

A: I input all information into Ecartico directly; for information that did not fit the Ecartico structure, I added references in Ecartico.

Q: Are you sharing your detailed research methods?

A: They will be published in my dissertation. I will also discuss with Harm Nijboer about making the structure of the research accessible. That will not necessarily make the research easily replicable, however: despite the availability of research methods, much of the research still consists in manual labour.

Q: Did you find data on female artists and patrons?

A: Ecartico includes the more important female artists and their pupils and servants; they were clearly present, but there is still a lot to discover. Judith Noorman's current research will enrich this.

Q: Did you do similar research for other cities?

A: I hope to do so in future research, but it takes a lot of time. It will be interesting not only for other cities as such, but also to study how intra-city and inter-city relationships functioned.

### ***Harm Nijboer (UvA) en Judith Brouwer (Huygens Instituut) – Onderzoek in een ogenblik. Prentkunst in Amsterdam in de zeventiende eeuw***

*Harm Nijboer is historicus en als datamanager en onderzoeker verbonden aan de Universiteit van Amsterdam. Samen met Judith Brouwer, Marten Jan Bok en Frans Blom voert hij de redactie over Ecartico. Hij geeft de presentatie mede namens Judith Brouwer, historisch letterkundige en cultuurhistoricus; zij werkt als datacurator en onderzoeksdatamanager bij het Huygens Instituut.*

Ecartico is te vinden op [www.vondel.humanities.uva.nl/ecartico/](http://www.vondel.humanities.uva.nl/ecartico/); deze pagina gaat binnenkort over in <https://ecartico.org>. Het is een databank met gegevens over bijna 60.000 schilders, graveurs en andere actoren in de culturele industrie in de 16<sup>e</sup>- en 17<sup>e</sup>-eeuwse Lage Landen.

Een databank is een gestructureerde verzameling gegevens, die het zoeken in gegevens makkelijker maakt (zoals een catalogus), maar het ook mogelijk maakt om verbanden tussen data te leggen en gegevens uit verschillende bronnen te aggregeren. Ecartico is begonnen om statistisch onderzoek te kunnen doen.

Een prentkunst-databank was snel opgezet. Voor prentkunst waren veel plaatjes en feitjes beschikbaar, maar weinig cijfers. Met deze databank kunnen nieuwe vragen worden gesteld en beantwoord: wie, wat, waar en hoeveel waren er van X? Of: wat was de relatie tussen prent- en boekproductie? Je kunt makkelijk een overzicht maken van bijvoorbeeld het aantal graveurs in Amsterdam in verschillende jaren; net als bij schilders zie je na 1670 een kleine terugval, maar anders dan bij schilders neemt het daarna weer toe en stabiliseert het zich op een hoger plafond. Ook kun je visualiseren waar ze geboren zijn. Velen komen uit Amsterdam en Nederland, maar graveurs zijn ook afkomstig uit de rest van Europa.

Je kunt prentkunstproductie zetten naast boekproductie. Daarvoor gebruiken we STCN (Short Title Catalogue Netherlands), beheerd door de KB; dit is een retrospectieve bibliografie van zo'n 250.000 boeken tot 1800. Er is enige discussie over de volledigheid hiervan, maar het is in elk geval een omvangrijke en nuttige bron. De titels zijn beschikbaar als Linked Open Data (RDF<sup>1</sup>), wat betekent dat je zelf queries kunt maken. Queries zijn zoekvragen in een bepaald formaat, zoals SPARQL,<sup>2</sup> die je kunt loslaten op een SPARQL-eindpunt, zoals <http://data.bibliotheken.nl/sparql>. Hoe ziet een query eruit? Dat lijkt moeilijk, maar valt best mee als je er even in verdiept en het een paar keer uitprobeert.

Je kunt bijvoorbeeld een zoekvraag maken naar het aantal boeken uitgegeven in Amsterdam in 1580-1710. Maar ook allerlei deelverzamelingen, bijvoorbeeld hoeveel boeken een gravure op het

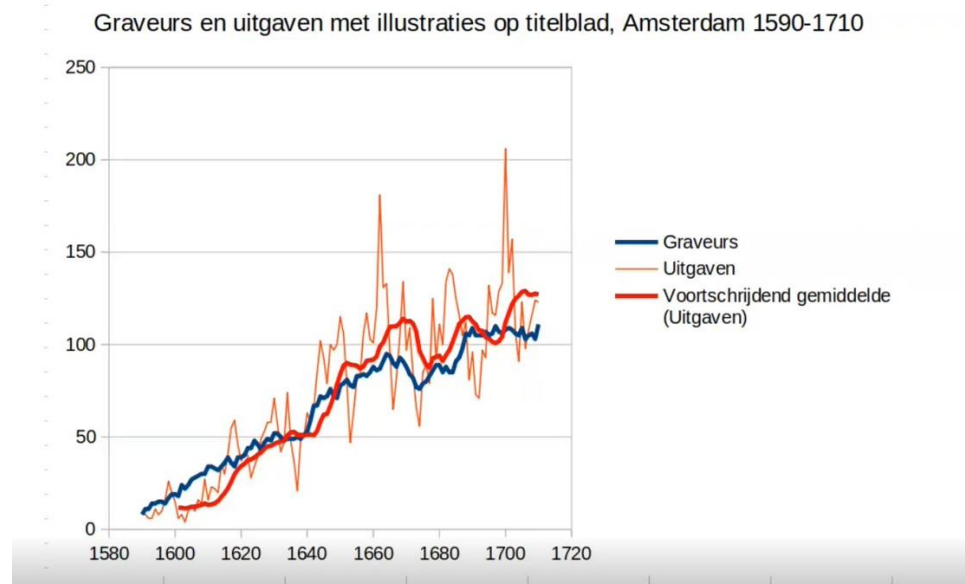
---

<sup>1</sup> Resource Description Framework, een (meta)datamodel, zie [https://nl.wikipedia.org/wiki/Resource\\_Description\\_Framework](https://nl.wikipedia.org/wiki/Resource_Description_Framework).

<sup>2</sup> SPARQL Protocol And RDF Query Language, een zoektaal om zoekvragen te maken in RDF-databestanden. Zie <https://nl.wikipedia.org/wiki/SPARQL>.

titelblad hebben. Deze statistieken kun je ook combineren, bijvoorbeeld het aantal boeken met een titelgravure en het aantal graveurs in Amsterdam. Die statistieken blijken – niet onverwacht – tamelijk parallel te lopen. Er blijkt zo een nauw verband te bestaan tussen de markt van boeken en de markt van prentkunst.

## Gegevens combineren



Een andere interessante invalshoek betreft familienetwerken. Deze mogelijkheid is niet opengesteld in Ecartico, omdat dit complexer is en de uitkomst niet makkelijk interpreteerbaar is. Je kunt bijvoorbeeld van alle 139 graveurs uit de periode 1670-1680 in Amsterdam de familieleden in kaart brengen; dan blijken er 2129 verwanten (tot de derde graad) te zijn, waaronder de nodige boekverkopers/drukkers (148) en schilders (203) zitten. Zo zie je dat prentkunst sterk samenhangt niet alleen met de schilderkunst maar ook met de boekdrukkunst. Op deze manier kun je netwerken en verbanden in kaart brengen.

Conclusie: dankzij (linked) open data kun je snel nieuwe resultaten ontwikkelen en zichtbaar maken. Je kunt ontzettend veel met data. Natuurlijk blijft ook (expert)kennis belangrijk, om zinvolle vragen te stellen, data goed te kunnen interpreteren en verbanden te vinden die zinvol zijn. Begrippen en ontologieën zijn daarbij belangrijk; begrip van begrippen is nog belangrijker.

Voor de toekomst is de visie: aggregatie van linked data, zie <https://www.goldenagents.org/>. Nu al te proberen: <https://sparql.goldenagents.org/>.

### Vragen

V: Hoe is Ecartico gefinancierd? Hoe lang duurde het om te bouwen?

A: Het is in meerdere fasen door middel van verschillende projecten gefinancierd. Eerst in een project van Eric Jan Sluiter en Marten Jan Bok, met data uit een kunstenaarslexicon dat Pieter Groenendijk had gemaakt. Daarna een databank over de hele Nederlanden, en nog andere projecten. Er wordt nog steeds aan gebouwd.

### Algemene vragen en discussie

V: Hoe kan de maker of eigenaar van een veranderd webadres een gebroken link ondervangen?

A: Je kunt het oude adres (de oude URL) in de lucht houden en dan nog lang een automatische doorverwijzing (*redirect*) daarop plaatsen.



Later komt de discussie nogmaals terug op 'linkrot', een probleem dat kennelijk breed speelt. Het helpt als instellingen persistente identificerende adressen (Persistent Identifiers<sup>3</sup>) hanteren, zoals *handles*, maar zelfs daarmee heb je geen garantie van permanente geldigheid. Ook *handles* moet je immers blijven onderhouden. Het probleem is dus nog niet opgelost.

V: wordt er ook gebruik gemaakt van Wikidata, en wat is Wikidata precies?

A: Wikidata bestaat 10 jaar. Het is een zusje van Wikipedia, alleen beschrijft het onderwerpen in een soort *linked data*-formaat. De gegevens kunnen dus door iedereen worden bevroegd; feitelijk is het een enorme databank waarin iedereen kan samenwerken. De meeste gebruikers zijn erfgoedprofessionals uit de praktijk. De kwaliteit is vaak best goed; Ecartico maakt er ook gebruik van. Je kunt personen uit beide databanken aan elkaar koppelen. Maar niet alle Ecartico-data gaan naar Wikidata, aangezien Wikidata een 'notability-criterium' hanteert: data moeten een zekere opmerkingswaarde hebben. Dit wordt bepaald door de Wiki-gemeenschap zelf en hoeft geen al te grote belemmering te zijn.

Wikidata verwijst ook naar andere relevante databanken, die onderzoekers niet altijd kennen (kijk bijvoorbeeld helemaal onderaan de pagina over de Nachtwacht). Blijf wel alert hoe betrouwbaar (correct en volledig) Wikidata is, omdat het minder systematisch is opgezet.

Er komen veel opmerkingen in de chat langs over Wikidata als belangrijke of nuttige bron. Esther Doornbusch (Rijksmuseum) kan tips geven. Daniëlle Jansen, linked-data-specialist bij de KB, vult aan: Wikidata is een datahub, waar je als erfgoedinstelling je eigen data kunt koppelen aan een heel groot netwerk. Daardoor kun je bijvoorbeeld een kaart maken van Nederland met alle straten die vernoemd zijn naar vrouwen uit de literaire vrouwendatabank van het Huygens Instituut. De KB is druk bezig STCN (de boektiteldatabank) te koppelen aan Wikidata, waardoor je geografische data ook makkelijk kunt visualiseren in kaarten.

Hans Nijboer voegt toe dat er wel beperkingen zijn, bijvoorbeeld als vrouwelijke verzamelaars niet in andere databanken staan; Wikidata heeft dan geen mogelijkheid om dit toe te voegen, dan zou je eerst een Wikipedia-pagina moeten maken over de desbetreffende verzamelaar. Maar, aldus Daniëlle Jansen, een vrouw hoeft maar in één bron ergens te worden genoemd om haar op Wikidata op te kunnen nemen.

Een interessant initiatief is *Sum of all paintings*: gebruikers proberen wereldwijd alle kunstenaars en alle 'notable paintings' op Wikidata te zetten.<sup>4</sup> Sommigen zijn bezig iets vergelijkbaars te doen voor verzamelaars.

V: Waar begin je als je iets wilt met Wikidata?

A: Wikidata heeft een SPARQL-endpoint. Wikidata kent bijvoorbeeld 'verzamelaar' als beroep, dus je kunt queries maken over verzamelaars en daar gegevens over opvragen.

## Deel II: Linked Open Data en Praktijk: Wat wordt er al gedaan en wat kunnen wij doen?

### **Reinier van 't Zelfde (RKD) – Door het LOD verbonden aan het RKD**

*Reinier van 't Zelfde is informatie-architect bij RKD. Hij ontwikkelde er de databank-infrastructuur, met de bijbehorende zoeksystemen, webportalen en webdiensten, die vandaag bekend staat als RKDExplore. Hij was betrokken bij projecten als de Rembrandt Database, Van Gogh Worldwide en The Mondrian Papers.*

<sup>3</sup> Zie [https://nl.wikipedia.org/wiki/Persistent\\_identifier](https://nl.wikipedia.org/wiki/Persistent_identifier).

<sup>4</sup> Zie [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_sum\\_of\\_all\\_paintings](https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings).

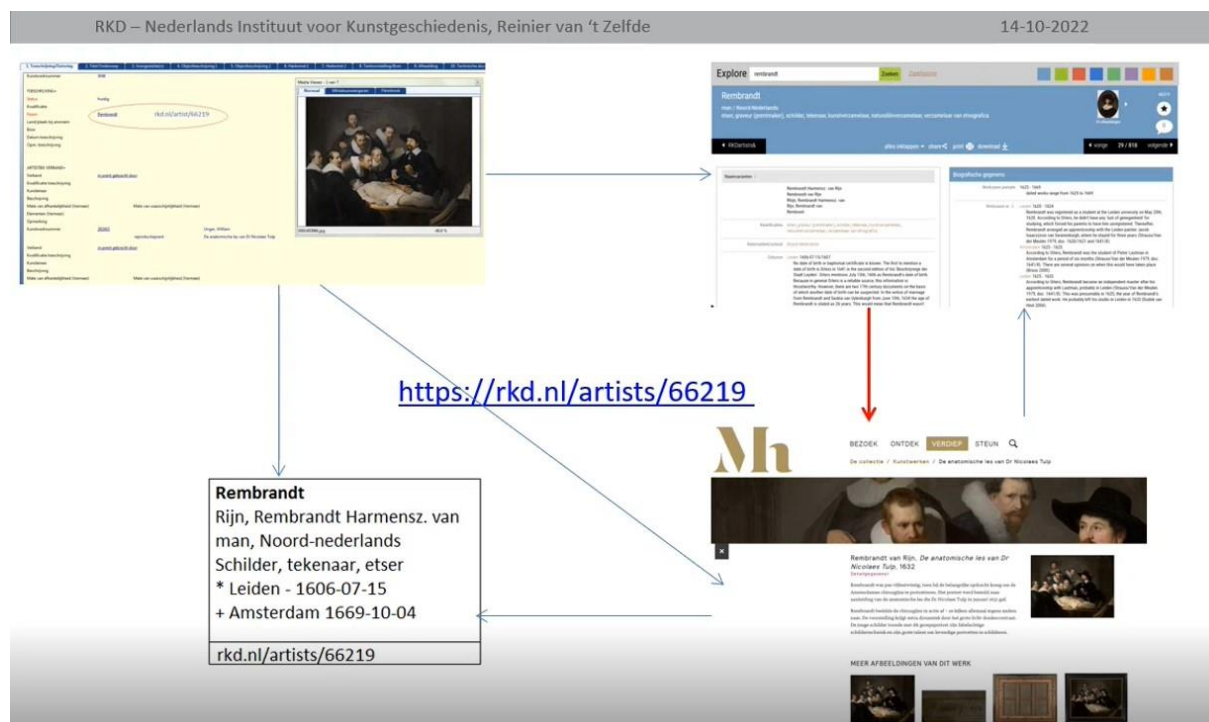
RKD's ambitie is om hét startpunt voor kunsthistorisch onderzoek naar Nederlandse kunst in een internationale context te zijn. De rol is verschoven van een *verzamelaar* van informatie naar een *verbinder* van informatie. Dat kan met Linked Open Data (LOD).

Er zitten zo'n 250.000 kunstenaars in RKDartists, die zijn allemaal verbonden in en met de andere RKD-databanken. Explore is het onderzoeksportaal ([www.rkd.nl](http://www.rkd.nl)). Er loopt een project om dit te vervangen door RKDresearch; alle databanken komen in deze omgeving als LOD beschikbaar. De beoogde opleveringsdatum daarvan is eind 2023.

LOD is gebaseerd op open standaarden (zoals CIDOC-crm, SKOS, RiCo en schema.org). Het is belangrijk om een passende ontologie<sup>5</sup> te gebruiken – je moet immers dezelfde taal spreken – bijvoorbeeld CIDOC-crm waarin museale objecten op een gestructureerde manier worden weergegeven.<sup>6</sup> Dit vergt ook eenzelfde type beschrijving van de inhoud, via standaardterminologie (bijv. AAT en Iconclass<sup>7</sup>) en standaarden voor de naamgeving van entiteiten, zoals ULAN<sup>8</sup> en vooral natuurlijk RKDartists.

RKDartists is generiek voor het kunst domein en bevat niet alleen kunstenaars maar ook verzamelaars, handelaren, kunsthistorici en critici. Ook belangrijke familieleden worden opgenomen, evenals geportretteerden – eigenlijk iedereen die voor het culturele werkveld van belang is. Elke pagina, met een eigen *persistent identifier* (een vaste URL), heeft een uitgebreide structuur van veel typen informatie, volgens gestandaardiseerde formaten. Dit is ook ontsloten via het Netwerk Digitaal Erfgoed (NDE).

De gegevens in RKDartists zijn dus als LOD beschikbaar. Dit biedt de mogelijkheid kunstenaars uit RKDartists te verbinden aan de eigen collectiebeheersystemen van musea. Ook kun je er gegevens geautomatiseerd aan ontlenu, zoals de voorkeursnaam, geboorte- en sterfdatum en -plaats en nationaliteit. Het Termennetwerk van NDE speelt daarbij een belangrijke rol.<sup>9</sup> Zo sluiten gegevens op



elkaar aan vanuit verschillende instellingen, en ook op andere databanken.

<sup>5</sup> Een ontologie is een formele beschrijving van elementen en de relaties daartussen in een bepaald domein, een soort combinatie van een typologie, een woordenboek en een thesaurus. Zie <https://www.ensie.nl/bibliotheecarium/ontologie>.

<sup>6</sup> Zie <https://www.cidoc-crm.org/>.

<sup>7</sup> Zie <https://www.getty.edu/research/tools/vocabularies/aat/> en <https://iconclass.org/>.

<sup>8</sup> Zie <https://www.getty.edu/research/tools/vocabularies/ulan/>.

<sup>9</sup> Zie <https://netwerkdigitaal erfgoed.nl/activiteiten/termennetwerk/>.



De data kunnen vrijelijk worden gebruikt op basis van een Creative Commons-licentie; wel wordt verzocht om de bron te vermelden met een link erbij. Overigens moeten de softwareleveranciers (zoals Axiell en Pictura) nog wel betere mechanismen implementeren om dit soort data goed te kunnen verwerken.

Het doel is uiteindelijk om in een netwerk samen te werken waarbij gegevens op elkaar aansluiten, in plaats van allemaal losse stukjes informatie te presenteren. Goede datakoppelingen kunnen ook dubbel werk bij instellingen voorkomen.

Tot slot het project Artists4all: nog niet alle relevante personen staan in RKD, veel biografieën moeten worden aangevuld, en fouten gecorrigeerd. Voor wat betreft vrouwen tussen 1780-1980 staan er al 5258 vrouwen in de databank, waarvan 472 verzamelaars. Maar er is nog steeds veel te doen. Het RKD biedt externe partijen die een serieuze (kwalitatieve en kwantitatieve) bijdrage willen leveren, toegang om dit te bewerkstelligen. Voorkom dubbel werk, help het RKD!

#### *Voor wie meer informatie of aan de slag wil*

- de informatie uit RKDartists kan in verschillende formaten worden weergegeven en geëxporteerd; zie meer informatie en toelichting hierover [link is vervallen]
- voorbeelden van verschillende output-formaten [links nog niet beschikbaar]
  - RDF
  - JSON/LD
  - Turtle
  - N/triples
- zie ook over datadumps in deze verschillende formaten:  
<https://data.netwerkdigitaalerfgoed.nl/rkd/rkdartists>
- Termennetwerk, validatieservice NDE: <http://demo.netwerkdigitaalerfgoed.nl/termennetwerk/nl>
- SPARQL-endpoint van NDE:  
<https://data.netwerkdigitaalerfgoed.nl/rkd/rkdartists/sparql/rkdartists>

#### **Vragen**

V: Hoe zit het met kwaliteitsborging van data en links?

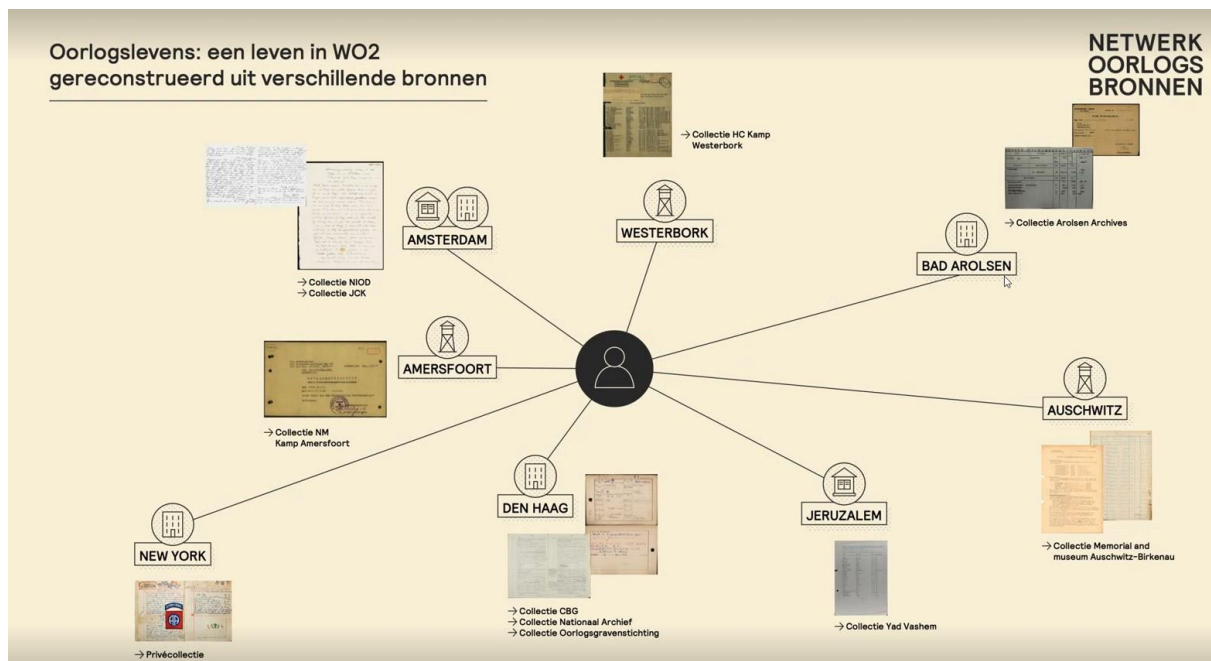
A: Het is de verantwoordelijkheid van RKD om alle links te laten bestaan; als blijkt dat namen dubbel zijn ingevoerd, blijft de ene wel bestaan maar wordt deze automatisch doorgelinkt. De betrouwbaarheid van informatie zelf hangt af van de bron (die meestal erbij wordt vermeld).

V: Maar hoe zit het met de duurzaamheid van links naar externe bronnen?

A: Dat is de verantwoordelijkheid van de bronhouder. Je zou dit geautomatiseerd periodiek kunnen controleren, maar dat gebeurt nog niet.

## Edwin Klijn (*Netwerk Oorlogsbronnen*) – Van archieven naar verhalen. 1 + 1 = 3

Edwin Klijn is sinds 2010 programmamanager van *Netwerk Oorlogsbronnen*. Hij is historicus en betrokken geweest bij verschillende digitaliseringsprojecten in de erfgoedsector.



Netwerk Oorlogsbronnen is een samenwerkingsverband van collecties over de Tweede Wereldoorlog (WOII), met 82 deelnemende instellingen. Het doel is de verhalen over WOII toekomst te geven door informatie te verbinden. Het is belangrijk om collecties aan elkaar te verbinden, om verhalen te kunnen vertellen.

Klijn illustreert dit aan de hand van Nathan Kornalijslijper. Over hem zijn verschillende gegevens beschikbaar uit uiteenlopende bronnen: rapporten (overgetikt door vrijwilligers) beschikbaar in Erfgoed Leiden (hij en zijn onderduikgever werden opgepakt), arrestantenkaarten in Stadsarchief Rotterdam (vertrek naar Vught; een kaart met Kornalijslijpers bezittingen), een kaart in de Arolsen Archives (Ned. Rode Kruis-archieef; hij wordt afgevoerd naar Westerbork), en kampadministratie (hij is vermoord in Auschwitz).

Ook zijn onderduikgever Theodorus Witte is traceerbaar in de verschillende bronnen. Witte werd tegelijk met Kornalijslijper verplaatst naar Vught en vervolgens als politiek gevangene tewerkgesteld in buitencommando's. Aangezien hij in Leids politierapport uit maart 1945 melding maakt van fietsdiefstal, is hij dan dus terug in Leiden.

Zo zijn er nog 299.998 verhalen, uitgezet op een automatisch gegenereerde tijdlijn. Bouwstenen van die verhalen zijn de gegevens (wie wat waar wanneer) van ruim 250 instellingen, met meer dan 400 collecties en ca. 60 persoonsdatasets. De verzameling wordt nog steeds uitgebreid. De meeste collecties zijn ontsloten via het Open Archives Initiative (OAI), sommige worden ook wel statisch aangeleverd, bijvoorbeeld via een zelfgemaakt Excel-bestand. Ze worden beter vindbaar gemaakt via één ingang: [www.oorlogsbronnen.nl](http://www.oorlogsbronnen.nl).

De data worden georganiseerd via *events modelling*: het clusteren van gegevens (plaats, persoon, thema) rond een gebeurtenis (bijv. iemands arrestatie, of de Februaristaking). Men gaat uit van *linked open data*, maar data worden vervolgens verrijkt (in een spinque-omgeving), mede via een WOII-thesaurus met contextuele informatie. Door de data te combineren, kan het héle verhaal worden verteld, wat meerwaarde heeft ten opzichte van de losse onderdelen: 1 + 1 = 3.

Daarbij verwijst het Netwerk Oorlogsbronnen ook terug naar de oorspronkelijke bronnen, zodat mensen ook de originele data kunnen raadplegen. Tot slot zijn de data ook beschikbaar als *linked open data* via het LOD-register, voor degenen die zelf ermee aan de slag willen. Dit register bestaat uit 92 datasets en is raadpleegbaar op <https://data.spinque.com/ld/data/netwerkoorlogsbronnen/>.

## Vragen

V: Hoeveel data was er al beschikbaar, hoeveel is nieuw (gedigitaliseerd)?

A: Collectiedata (beschrijvingen) waren beschikbaar. Voor wat betreft persoonsdata hadden een paar grote instellingen al gegevens beschikbaar, kleinere niet. Maar heel veel wordt nu systematisch via *crowdsourcing* ontsloten, ook bij de meeste regionale archieven.

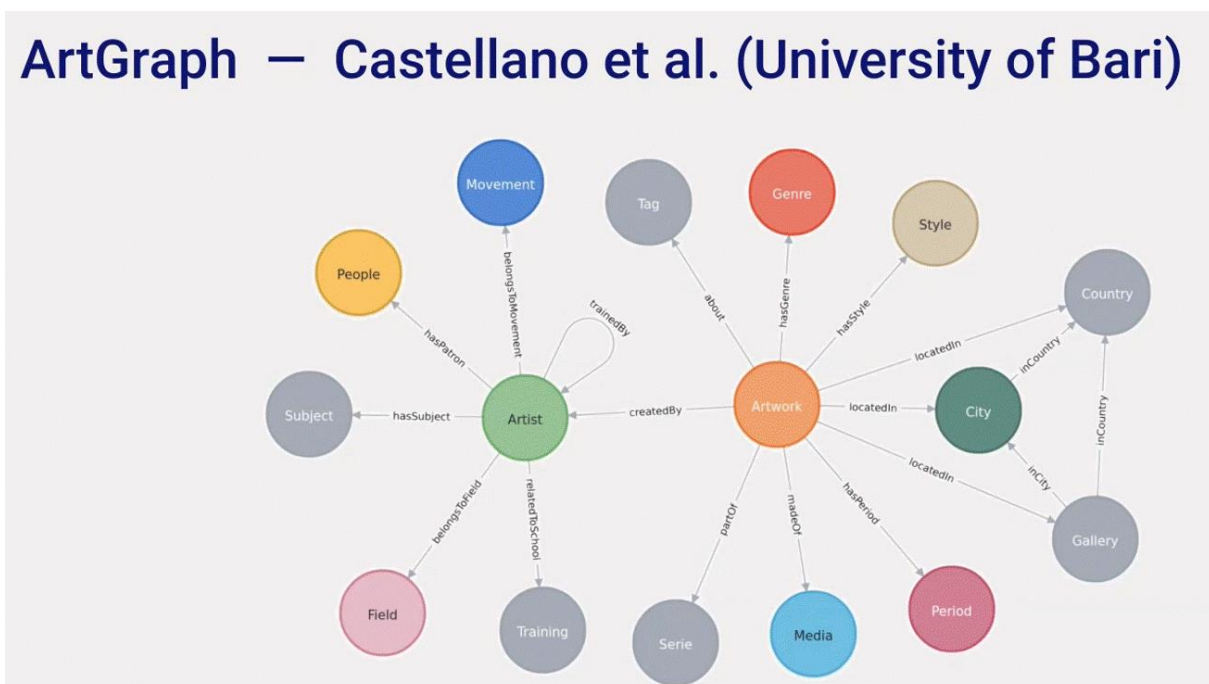
## Daniel Kapitan (EASI) – Knowledge Graphs in de praktijk

*Daniel Kapitan, gepromoveerd fysicus, is adviseur digitalisering, datawetenschappen, strategie en organisatieverandering. Hij doceert aan de Jheronimus Academy of Data Science en is Fellow bij het Eindhoven AI Systems Institute (EASI). Hij probeert vooral organisaties met een publieke taak te helpen meer te doen met data.*

Kapitan werkt aan het breder toepasbaar maken van nieuwe technologieën, zoals *knowledge graphs* (kennisgrafen) en geautomatiseerde data-analyse met behoud van privacy. Kennisgrafen zijn al heel oud, zie het 18<sup>e</sup>-eeuwse bruggenprobleem van Leonhard Euler. Het gaat om het modelleren van knooppunten (*nodes*: entiteiten met kenmerken) en de relaties daartussen. Kennisgrafen en *linked data* zijn gebaseerd op dezelfde principes. Alle technische afkortingen die in eerdere presentaties zijn genoemd, zijn simpelweg technische implementaties van deze principes.

Databanken zijn ingeburgerd en worden veel gebruikt, maar ze zijn inherent beperkt: als je nieuwe relaties tussen data in een databank wil leggen, moet je een nieuwe tabel maken (bijvoorbeeld een nieuw Excel-blad). Daarom zijn andere structuren ontwikkeld, zoals *document stores*, waarin *key value pairs* worden gedefinieerd. Maar bij de huidige hoeveelheden data blijft dat moeilijk visualiseerbaar. Daarom zijn velen bezig databanken te integreren in kennisgrafen, waarin je semantisch kan zoeken (op taal), niet alleen op termen maar ook op inhoud of onderliggende relaties.

Een voorbeeld is ArtGraph,<sup>10</sup> waarmee je bijvoorbeeld clusters van kunstenaars kunt vinden, of genres kunt ontdekken die nog niet (duidelijk) waren gedefinieerd. Dit opent de deur voor nieuwe analysetechnieken, bijvoorbeeld het linken van personen via ‘vriend van een vriend’, waarbij je de



<sup>10</sup> Zie Giovanna Castellano et al. (Bari), “Leveraging Knowledge Graphs and Deep Learning for automatic art analysis,” *Knowledge-Based Systems* 248 (2022), <https://doi.org/10.1016/j.knsys.2022.108859>.

resultaten dan begrijpelijk kunt visualiseren, ook bij grote aantallen. Veel is dus mogelijk, mits je data eerst goed digitaliseert, dan annoteert met geharmoniseerde woordenboeken, en vervolgens ontsluit.

Er zijn ook interessante voorbeelden uit andere sectoren, zoals de Stelselcatalogus: de integratie van de ca. 10 basisregistraties in Nederland, zoals personen, adressen en gebouwen, percelen en voertuigen. Hierdoor kan bijvoorbeeld de Belastingdienst vooraf ingevulde formulieren sturen. Koppeling roept natuurlijk privacyragen op, zodat er technieken worden ontwikkeld om privacy-bestendig te koppelen. Los daarvan is koppeling ingewikkeld, want alleen al een huis staat in vier verschillende basisregistraties op verschillende manieren geregistreerd. Je moet dezelfde entiteit (het huis) dus heel gedetailleerd uitwerken, wil je data kunnen combineren.

Een ander voorbeeld is Sensing clues: het tegengaan van stroperij in beschermde parken in Afrika (<https://sensingclues.org>). Hierbij wordt allerlei soorten informatie van verschillende partijen uit het veld *real-time* vergaard. Om de verschillende typen gegevens met elkaar in verband te brengen, kun je het beste kennisgrafen gebruiken, op basis van geautomatiseerde tekstanalyse. Daarbij worden teksten samengevat via subject-predicaat-object-trio's (Tom Sawyer – ontmoet – een vriend; Tom – is – blij); dit maakt het mogelijk om krachtiger samen te vatten.

Er zijn nog veel meer voorbeelden, zoals KIK-V (informatie-uitwisseling in verpleeg- en verzorgingstehuizen en thuiszorg) en Wendbaar wetgeven (feitmodellering als basis voor een betrouwbare digitale overheid).

Een belangrijke ontwikkeling is FAIR Digital Objects (FDO),<sup>11</sup> waarbij FAIR uitdrukt dat data vindbaar, toegankelijk, interoperabel en herbruikbaar (Findable Accessible Interoperable Reusable) moeten zijn. FDO is eigenlijk het al langer bestaande toekomstvisioen van het semantische web: een wereldwijd web voor alle data. Het duurt nog wel even om te realiseren.

Kapitans heeft tot slot een suggestie voor het project *De Andere Helft*: maak werk van betrouwbare data. Denk goed na hoe je dingen vastlegt, in de juiste context, op een standaardmanier. Betrouwbare data zijn data die geschoond zijn, gegrond in gedeelde begrippen, gecontextualiseerd en toegankelijk in een gestandaardiseerd formaat, en die worden beheerd en geactualiseerd.<sup>12</sup>

## Vragen

Geen vraag maar een aanvulling: bij de Digitale Erfgoed Referentie Architectuur, die vanuit het ministerie wordt ontwikkeld, wordt instellingen ook geadviseerd data volgens de FAIR-principes te ontsluiten.

## Algemene vragen en discussie

V: AdLib zal op korte termijn minder worden gebruikt wegens onderhoud (en gaat over in Axiell Collections, een webgebaseerde variant). Is dit niet hét moment om de beheersystemen van alle instellingen af te stemmen om op hetzelfde registratiesysteem uit te komen? Zijn er significante verschillen tussen TMS en Axiell in hoe zij data vastleggen?

A: Het beheersysteem zelf maakt niet zoveel uit, het gaat erom dat je data op een gestandaardiseerde manier vastlegt. Het is daarvoor belangrijker om het formaat van geregistreerde data af te stemmen, en dus ontologieën te gebruiken. Musea voeren gegevens al grotendeels in op basis van thesauri.. Het probleem lijkt dan meer te zitten in het delen van informatie.

A: Kijk vooral ook naar de uitgangspunten van Common Ground.<sup>13</sup> Zie data los van de applicatie waarin ze toevallig ooit zijn gegoten. Dit klinkt als een open deur, maar is het niet. Zie bijvoorbeeld de zorgsector waarin sommige instellingen weigeren zich te conformeren aan nieuwe standaarden.

---

<sup>11</sup> Zie <https://fairdo.org>.

<sup>12</sup> Zie George Fletcher, Paul Groth en Juan Sequeda, “Knowledge Scientists: Unlocking the data-driven organization” (2022), <https://arxiv.org/abs/2004.07917>.

<sup>13</sup> Zie <https://vng.nl/artikelen/common-ground>.

Je voorkeuren mogen interoperabiliteit niet in de weg zitten. Het moet niet uitmaken of je TMS of Axiell Collections gebruikt, zolang je maar standaarden gebruikt.

V: Hoe kunnen we in of vanuit *De Andere Helft* beginnen met delen van informatie?

A: Het lijkt ingewikkeld om de techniek van kennisgrafen aan te leren, maar het is te doen, en je kan er veel meer mee, dus het loont de investering.

A: Maak het niet te exotisch, belangrijk is dat we data goed uitwisselbaar maken. Kijk waar je op kunt aansluiten.

A: Begin klein. Probeer niet de hele wereld te omvatten, maar breid gaandeweg uit, al lerend doende. Een kleine relationele databank kan al zinvol zijn als startpunt. 'Klein' betekent, in grafen gedacht, dat je een beperkt aantal eigenschappen opneemt (bijvoorbeeld geboorte- en sterfdatum, plaats, type kunstenaar). Zorg overigens vooral dat je aansluit bij bestaande databanken, als je data gaat verzamelen en toevoegen.

Overigens vergt onderzoek soms andere eigenschappen dan documentatiedoeleinden, dus het hangt af van wat je met de data wil.

V: In dat licht, zijn er al onderzoekers die hun data kunnen benoemen en iets kunnen zeggen over mogelijkheden voor hun datasets?

A: Bram Donders heeft wel vrouwelijke kunstenaars/verzamelaars geïdentificeerd en gekoppeld aan geboorte/sterf/huwelijks-data en kunstwerken, maar dat zit nog niet in een databank met *linked open data*. Hij wil daar wel naar toe, zodat bijvoorbeeld van alle vrouwelijke leerlingen van Allebé kan worden geanalyseerd wie er bijvoorbeeld naar Oosterbeek of Dongen trok. Hij wil dat ook visualiseren via een graaf.

Dat roept wel de vraag op hoe de andere andere helft – mannen – zich verhoudt tot vrouwen: ze behoren immers gezamenlijk tot netwerken. Je kunt de andere helft niet los onderzoeken; ook de ene helft moet dus bekend zijn. Maar mannelijke kunstenaars en verzamelaars zitten in principe al in RKDartists, dus het gaat mogelijk worden om vergelijkingen te maken en verbanden te leggen. Vooralsnog is *De Andere Helft* vooral bezig met inventariseren en verankeren, daarna volgen de onderzoeksvragen. Eerst de bouwstenen, dan het bouwwerk. Niettemin bestaan veel bouwstenen al: informatie in de systemen van alle instellingen, bijvoorbeeld van (vrouwelijke) schenkers. Door die informatie te koppelen, kun je veel informatie naar boven krijgen die nu verborgen blijft. Idealiter gebeurt dat via *linked open data*.

Belangrijk is vervolgens dat door koppeling gevonden nieuwe informatie wordt vastgelegd in bij voorkeur bestaande databanken zoals RKDartists, zodat die informatie gelijk ook vindbaar wordt en gebruikers niet in weer een nieuwe databank moeten gaan zoeken. Het mooie van Netwerk Oorlogsbronnen is dat het zelf geen data cureert, maar data koppelt en visualiseert, waarbij de gebruiker de data zelf kan (en moet) interpreteren. Bijvoorbeeld bij tegenstrijdige bronnen wordt geen standpunt ingenomen welke bron correct(er) is. Met de verrijkingsslag voegt Netwerk Oorlogsbronnen overigens wel data toe, in een extra laag – een vorm van annotaties bij de "ruwe" data. De scheiding tussen brongegeven en annotatie is belangrijk: historici maken immers een scherp onderscheid tussen brongegeven en interpretatie.

## Afsluiting

Mayken Jonkman bedankt de sprekers namens alle deelnemers voor een briljante ochtend.

Toekomstige activiteiten in relatie tot *De Andere Helft*:

- 7 maart 2023: groot congres in het Rijksmuseum, *Women in the Museum, Glass Ceilings*;
- 1-2 juni 2023: ESNA-congres over vrouwen in de context van hun tijd en plaats.

Volg de RKD-webpagina. En geef activiteiten rond vrouwen in de kunst door aan RKD.