Van Mourik Broekmanweg 6
2628 XE Delft
Postbus 49
2600 AA Delft

www.tno.nl

T +31 88 866 30 00
F +31 88 866 30 10

TNO-rapport

TNO 2014 R11049 | Eindrapport

Kansen voor Big data – WPA Vertrouwen

Datum	16 juli 2014
Auteur(s)	Tijs van den Broek, Arnold Roosendaal, Anne Fleur van Veenstra en Anna van Nunen
Exemplaarnummer	
Oplage	
Aantal pagina's	49 (incl. bijlagen)
Aantal bijlagen	2
Opdrachtgever	Samenwerkingsmiddelen onderzoek (SMO)
Projectnaam	Kansen voor Big data – WPA Vertrouwen
Projectnummer	060.08321

Summary

Big data is expected to become a driver for economic growth, but this can only be achieved when services based on (big) data are accepted by citizens and consumers. In a recent policy brief, the Cabinet Office mentions trust as one of the three pillars (the others being transparency and control) for ePrivacy. As such, it is a requirement for realizing economic value of services based on (personal) data. Businesses play a role in guaranteeing data security and privacy of data subjects, but also government organizations may facilitate these developments, for example by creating regulation, or by setting standards. Often big data services are based on data sharing among different organizations. This may impact trust in these services. Furthermore, organizations can become responsible for any privacy breaches by organizations they collaborate with. This risk can hamper creating new data based services by multiple organizations. Therefore, this project investigates how governance of big data services in settings of multiple organizations is organized and which requirements are necessary for creating trust in these services.

Business model of big data: a need for clear ownership and accountability
The business model of big data is based on data maximization: when more datasets are available and combined to create new data, the value of these data increases. This may, however, clash with privacy and data protection. The Dutch Data protection act encompasses particular purpose limitation, which holds that personal data may only be used for the purpose for which they are collected. Big data may also lead to the creation of personal data, when data are combined in such a way that they again form data that can be traced back to individuals. Therefore, to ensure trust in big data applications, data ownership and accountability for privacy purposes need to be made clear. Data ownership is important in creating trust and sufficient data quality '**ex ante**'. Accountability makes clear where data originates and in which way the privacy of individuals is guaranteed while data are processed ('**ex post**'). To allow for accountability and tracing data through audits, transparency needs to be in place.

This study investigates trust in big data services based on four aspects: form of collaboration, privacy, data ownership and accountability for privacy. The main research question is: *"How to ensure data ownership and accountability in order to minimize privacy risks in collaborations for big data services?"* This research question is addressed in three steps. Firstly, desk research is carried out to determine characteristics of big data in relation to privacy risks, to determine the role of data ownership and accountability in ensuring privacy of big data applications, and to identify different forms of collaboration in networks. Secondly, four use cases are investigated to find out how these elements are implemented in practice and subsequently, an analysis of these use cases is carried out. Thirdly, conclusions and a discussion of the findings are formulated.

Characteristics of big data in relation to privacy
The characteristics of big data and their relation to privacy are investigated by looking at three stages of data processing: collection, analysis, and application. During the data collection phase the main question is whether collecting the data is allowed. Data protection regulation presents requirements for specific purpose binding, information obligation to data subjects and organizational and technical

measures. Regarding big data these requirements may present problems, as the purpose of data processing may not be clear at first and may therefore differ from the purpose for which data are collected. In the analysis phase datasets may be combined. This may have two different implications: it could lead to re-identification of data to individuals, or it could lead to greater anonymity as the dataset grows. Therefore, in this phase, the ban on automated decision making is of importance.

The application phase may have the strongest implications for individuals or for groups. Profiling and predictions often represent an average based on queries on a database, and thereby they are often not applicable to all persons. However, current regulation appears to mainly apply to data collection rather than to data application. Therefore, a gap appears between the law and the technological practice. Also when data cannot be traced back to individuals, they can have substantial impact on persons when profiling is applied to categories of persons. Thus, while most impact occurs on the application stage, this is where privacy regulation covers least. Finally, big data may not comply with the notion of data minimization, which is also part of the data protection act, as much of the value of big data can be found in secondary applications of data that are not foreseen upfront.

Data ownership and accountability as protecting privacy in collaborations
As current regulation does not naturally comply with the premises of big data, it is interesting to see how additional policy measures can be used to guarantee the privacy of individuals. Two of those measures are data ownership and accountability. Ownership concerns the rights and obligations that organizations or individuals have regarding specific datasets. Accountability relates to the responsibility of organizations to account for the way in which data is processed. The way in which this notion is used in this study relates to ways in which activities are accounted for, as well as the ways in which data are collected and applied. As this also relates to any consequences of the use of data and ensuring proper requirements and remedies, accountability in this study refers to more than merely attributing any irregularities in data. To assure accountability, transparency is a requirement as the origin of data needs to be traceable, for example via audits.

The way in which organizations collaborate is essential for organizing accountability. Generally, four types of collaboration can be distinguished, depending on the type of coordination that is dominant. In a *market* the autonomy of organizations is central and collaborations happen ad hoc via market transactions. In a *bazar* reputation and community are central notions, as organizations independently create products and services and transactions take place ad hoc. In a *hierarchy* formal power relations are central and control takes place via sanctions or rewards. A *network* refers to a type of collaboration in which participants jointly coordinate activities, decisions, distributions of means and conflicts, and where trust is the main coordination mechanism. The type of collaboration depends on a number of factors. When applying the types of collaborations to data sharing for creating big data applications, different constellations emerge, based on the type of transaction, data ownership, and accountability.

In a market or bazar a dyadic *transaction* is central: buying or selling data, while in a hierarchical relation or network data sharing is structural. In a network or market, *ownership* of data remains at the organizations where the data originates, while agreements on the use and sharing of data are made via licenses. In a hierarchical

type of collaboration, a central organization owns the data and in a bazar, organizations refrain from owning data using a creative commons license. *Accountability* for data collection, analysis and application in a market or bazar resides at individual organizations. In a hierarchy, this resides with the central organization, while in a network type of collaboration, accountability needs to be jointly determined. As organizations will always try to limit privacy risks by increasing control, it appears that types of collaboration based on long-term relations are more desirable. This is even more the case for collaborations in which sensitive information is shared. However, increased control makes peer-to-peer collaboration harder to establish, while this may be more desirable from the point of view of user empowerment.

Use cases and cross-case analysis
Based on theoretical sampling four use cases are investigated in this study: Ahold personal marketing, Rotterdam open data, Achmea Health Database and energy data sharing. Ahold personal marketing concerns the processing of data collected via its loyalty program for marketing purposes. While this case was chosen on the idea that it would represent a market model, based on the empirical investigation, it appeared that in fact a hierarchical collaboration was present. The Rotterdam open data portal presents a bazar set up around open datasets from the municipality of Rotterdam. The hierarchical collaboration around the Achmea Health Database, an epidemiological dataset in which all data on the use of healthcare of those that are insured via the insurance company are collected for administrative purposes. These data can be reused for scientific purposes. Energy data sharing aims to establish a network type of collaboration around a platform for the use of smart meter data for different purposes.

In line with the description of the use cases, the cross-case analysis concerns four aspects: big data collaboration, characteristics of big data, privacy risks, and ownership and accountability. Regarding *big data collaboration*, none of the cases were found to represent the market governance form, which means that in none of the cases an example could be found in which data was shared openly for a commercial purpose. All cases indicated that they could not yet establish a business case for sharing data in this manner. Research and innovation were the most often found reasons for sharing data. The cases varied from 1-to-1 data sharing (Ahold personal marketing), to 1-to-many (Rotterdam open data en Achmea Health Database), to many-to-many (energy data). The 1-to-1 model offers organizations most control; the many-to-many model is most complex. Complexity determines the openness of collaboration. While the Ahold and Achmea cases represent closed models, Rotterdam and the energy data represent an open form of collaboration. Regarding the *big data characteristics*, sharing and combining data does not take place on a large scale, which also means that few collaborations take place at the moment. All cases predominantly share structured data such as transaction data, rather than unstructured data such as social media data. Furthermore, few cases show sign of data maximization, which means that the potential of big data to come up with unpredictable applications is not yet realized. However, all cases expect that the use of data will increase in the future.

All cases show strong awareness of the *privacy risks* involved in sharing data, which is often a result of the existence of the data protection act (Wbp). This means that the cases often focus on compliance and in at least one case (Rotterdam open data) this means that no data is shared that contains personal data. The cases

further show that obtaining (informed) consent from individuals is often difficult. The Achmea Health Database case even stated that the costs of obtaining consent are expected to be higher than the revenues. Data use for scientific purposes does take place, but initiatives of data processing for more general societal purposes are not yet observed. Furthermore, consent is usually obtained by having people accept general terms, which is not very elegant, nor does it have a strong legal basis. Contrary to our expectations the cases do not show that specific purpose binding is considered a problem. An explanation may be that organizations in all use cases are very careful in applying big data, which means that they are also careful in determining the purposes for data processing before asking consent. All cases hold that there are still many uncertainties involved in sharing data within a network of organizations.

Ownership is often still unclear in the use cases, also regarding personal data. In other cases, ownership of the data by the organization that possesses the data is assumed. Most questions regarding ownership arise in the field of medical care, where sensitive personal data are processed. Questions regarding ownership arise on two levels. Firstly, between organizations that share data and, secondly, between organizations and individuals. The notion of user empowerment is raised in all cases. While user empowerment is expected to increase trust in data services, it is also expected to weaken the business case of big data initiatives. All cases raise the question of ownership of combined data. In practice, this is still agreed on an ad hoc basis. *Accountability* is realized by some organizations from the perspective of compliance. This means that internal and external audits are undertaken. This may also be realized by maintaining a hierarchical relation, which means that control can be exerted over data and how it is used. Another means of realizing accountability is setting up shared facilities such as Trusted Third Parties (TTPs). In open forms of collaboration, accountability leans towards realizing transparency, information supply and user empowerment to allow individuals to control their data. Especially in networks also trust plays a large role besides agreeing on accountability.

Conclusions and discussion

Organizations aiming to share data need to determine how their collaboration will take place, which data they will share, which privacy risks are involved, and how ownership and accountability can address these risks and increase trust, in addition to ensuring compliance with the data protection act. This study shows that two barriers are in place for data collaborations: finding a business case and the existence of strict legislation that may not always support developing big data applications. Besides compliance to the data protection act, determining ownership and accountability may increase trust in data collaborations. This could be realized on four levels: process, ownership, dataset, and algorithm. To ensure trust, it may become more and more important to create applications in which individuals are in control.

Inhoudsopgave

1 Inleiding

1.1 Context

Vertrouwen is een van de drie pijlers die genoemd zijn in de kabinetsvisie ePrivacy[1] (naast transparantie en controle), en vormt een belangrijke voorwaarde voor het realiseren van economische waarde naarmate de rol van (persoonlijke) data in steeds meer sectoren steeds groter wordt. Om de economische waarde uit data te vergroten, is het dan ook van belang om het vertrouwen in data-toepassingen te vergroten. Bedrijven spelen hier een belangrijke rol in, om het vertrouwen van consumenten te wekken, maar mogelijk kan ook de overheid hier een rol in spelen. Bijvoorbeeld door wet- en regelgeving aan te scherpen, of om het opstellen van kaders of richtlijnen te stimuleren. Dit werkpakket heeft dan ook als doel om in kaart te brengen welke randvoorwaarden aanwezig moeten zijn om dit vertrouwen te wekken in big data toepassingen.

Zowel wanneer big data oplossingen en diensten worden ingericht als wanneer data diensten worden geleverd moet dit vertrouwen er zijn. Tijdens het inrichten van governance moet er bijvoorbeeld gezorgd worden dat partijen elkaars data gaan gebruiken en dat de privacy van individuen beschermd blijft. Wanneer oplossingen zijn ingericht, geldt dat er duidelijkheid moet zijn over de oorsprong en kwaliteit van data en dat, indien er een privacyschending heeft plaatsgevonden, er voldoende waarborgen zijn. De eigenschappen van big data verhogen echter de kans op privacyschendingen. Big data staat voor het gebruik van datasets die te groot zijn om met reguliere IT toepassingen te verwerken. Big data wordt gekarakteriseerd door een groot volume, hoge snelheid en grote variatie in data. Het koppelen van datasets leidt weer tot nieuwe data. Het business model van big data gaat uit van datamaximalisatie: hoe groter de datasets en hoe meer datasets met elkaar gekoppeld worden, hoe hoger de potentiële waarde (bijvoorbeeld nieuwe toepassingen). Het big data business model is echter niet zonder risico's. Het uitgangspunt van datamaximalisatie wringt namelijk met de bescherming van persoonsgegevens. De Wbp vergt namelijk doelbinding: data met persoonsgegevens mogen alleen gebruikt worden voor het doel waarvoor ze zijn verzameld. Daarnaast kan het combineren van datasets leiden tot (indirecte) persoonsgegevens bij de verwerking en toepassing van de data. De bewuste onvoorzienbaarheid van de toepassingen van big data kunnen dus leiden tot onvoorziene privacyrisico's.

Om te zorgen voor vertrouwen in big data oplossingen is het noodzakelijk dat waarborgen als eigenaarschap en accountability zijn geregeld. Eigenaarschap van data speelt een belangrijke rol bij het creëren van vertrouwen in data en datakwaliteit '**op voorhand'**. Bij eigenaarschap speelt dat degene die over de data beschikt, mogelijk een databankenrecht, auteursrecht, of bepaalde licenties heeft. Diegene heeft dus bepaalde rechten/aanspraken en bijbehorende verantwoordelijkheden. Accountability speelt een belangrijke rol bij het duidelijk maken waar data vandaan komt en op welke wijze de privacy gewaarborgd wordt

[1] Ministerie van Economische Zaken (2013). Kabinetsvisie op e-privacy: op weg naar gerechtvaardigd vertrouwen, http://www.rijksoverheid.nl/documenten-en-publicaties/kamerstukken/2013/05/24/kamerbrief-met-kabinetsvisie-op-e-privacy.html.

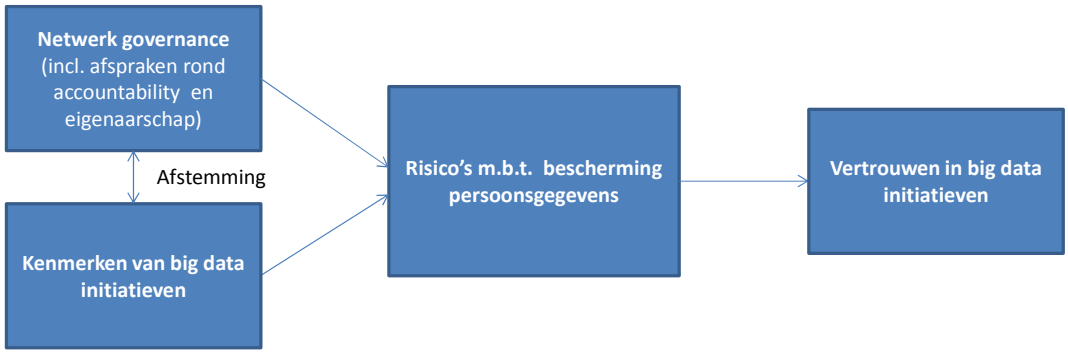
gedurende het verwerkingsproces en ook '**achteraf**'. Voor accountability is het noodzakelijk dat er transparantie is over waar data vandaan komt en dat het mogelijk is om dit na te gaan bijvoorbeeld via audits. Daarnaast speelt ook nog dat data verschillend van aard kan zijn; er zijn gegevens die bewust door individuen zijn verstrekt, er zijn metadata over klikgedrag en er is afgeleide informatie doordat datasets worden gecombineerd en profielen worden opgesteld. Al deze zaken zijn van invloed voor het inrichten van big data oplossingen, die ook nog verschillen in bepaalde situaties, zoals B2C en B2B[2].

Daarnaast overschrijdt big data vaak organisatiegrenzen: data over logistieke processen bestrijkt bijvoorbeeld een gehele keten waarin meerdere leveranciers samenwerken. De data van partners kan een competitief voordeel geven, bijvoorbeeld door efficiënter of effectiever bedrijfsprocessen in te richten. Daarnaast is er specialistische kennis en een kostbare ICT infrastructuur nodig om big data waardevol in te zetten. Daarom werken steeds meer organisaties samen op het gebied van big data. Deze samenwerkingen zijn niet zonder risico's voor de privacy. Als eigenaarschap en accountability onvoldoende zijn ingericht, dan zijn netwerkpartijen niet alleen verantwoordelijk voor privacyschendingen in hun eigen organisatie, maar ook voor de privacyschendingen door partijen waarmee ze samenwerken. Naast juridische risico's kan een gebrek aan afspraken leiden tot een afbreukrisico als samenwerkende partijen privacywetgeving schenden. Door deze verhoogde risico's is het belangrijk dat eigenaarschap en accountability afgestemd worden tussen de samenwerkende partijen.

In WPA wordt vertrouwen dan ook ingevuld aan de hand van vier aspecten: samenwerkingsvorm, privacy, eigenaarschap en accountability.

1.2 Probleemstelling

Dit werkpakket is er op gericht om duidelijk te krijgen hoe bij het samenwerken op het gebied van big data accountability en eigenaarschap kunnen worden ingericht om privacyrisico's te minimaliseren.



Figuur 1 Onderzoeksmodel

Het bovenstaande onderzoeksmodel laat zien welke relaties in dit onderzoek worden bestudeerd. De hoofdonderzoeksvraag is:

[2] Dit speelt ook bij G2C en B2G situaties, maar deze worden in dit werkpakket niet meegenomen.

"Hoe richt je eigenaarschap en accountability in om privacyrisico's bij het samenwerken met big data te minimaliseren?"

De onderzoeksvraag is onderverdeeld in de volgende deelvragen:
1. Wat zijn de eigenschappen van big data en hoe leiden deze tot privacyrisico's?
2. Wat is de rol van eigenaarschap en accountability als waarborgen van privacy bij het verwerken en toepassen van big data?
3. Welke samenwerkingsvormen zijn te onderscheiden voor het samenwerken rondom big data?
4. Hoe kunnen afspraken over accountability en eigenaarschap in verschillende samenwerkingsvormen helpen om privacyrisico's te minimaliseren?

1.3 Aanpak

Om deze onderzoeksvragen te beantwoorden, bestaat deze studie uit drie stappen:
1. Desk research naar eigenschappen van big data, privacy, eigenaarschap en accountability en samenwerkingsvormen om randvoorwaarden te identificeren. Bij de randvoorwaarden en aspecten waaraan voldaan moet worden, wordt onderscheid gemaakt tussen die zaken die op dit moment wettelijk al goed geregeld zijn, tussen die zaken die geregeld moeten worden in een commerciële relatie en vastgelegd kunnen worden in contracten tussen partijen en een 'grijs' gebied van zaken die (nog) niet in van beide categorieën vallen. Het desk research moet een framework opleveren dat gebruikt wordt bij het case study onderzoek en dat gevalideerd en/of aangevuld wordt in dit empirische onderzoek.
2. Opstellen van use cases over vier praktijksituaties van big data om te bekijken hoe de aspecten en randvoorwaarden zijn ingericht dan wel moeten worden ingericht. Om de situaties voldoende te laten verschillen, wordt er gekozen voor use cases die verschillen in samenwerkingsvorm, en dus in de controle over verzamelen, analyseren en toepassen van data.
3. Het analyseren van de use cases, waarbij ook een vergelijking wordt gemaakt tussen de aspecten en randvoorwaarden die zijn ingericht. Doel van deze stap is om inzicht te krijgen in de verschillende aspecten die moeten worden geregeld voor het inrichten van vertrouwen bij het samenwerken op het gebied van big data oplossingen.

De use cases worden opgesteld aan de hand van desk research en interviews met betrokken partijen. Voor de selectie van de use cases is gebruik gemaakt van theoretical sampling, waardoor er gekozen is voor use cases die verschillende typen samenwerkingsvormen vertegenwoordigen. Er is nadrukkelijk gestreefd om per use case verschillende partijen te spreken, bijvoorbeeld een consument, de dataverzamelaar en de datagebruiker. Dit kunnen overigens ook rollen zijn die op verschillende plekken in een organisatie zijn ingericht.

In het volgende hoofdstuk worden de bevindingen van het literatuuronderzoek gepresenteerd. Vervolgens worden in hoofdstuk 3 de use cases beschreven. In hoofdstuk 4 worden de bevindingen van de analyse van de use cases gepresenteerd, gevolgd door conclusies en discussie in hoofdstuk 5 en 6.

2 Theorie

In het desk research brengen we in kaart welke aspecten een specifieke rol hebben in de relatie tussen samenwerkingsvorm, privacy, eigenaarschap en accountability. Vanuit economisch perspectief is interessant om te bezien wat een gezonde balans is tussen het gebruik van data, waaronder persoonsgegevens, en de economische kansen die daaruit voortvloeien, en de bescherming van de privacy van degenen om wiens gegevens het gaat. Het literatuuroverzicht bestaat dus uit vier delen: de eigenschappen van big data en de privacyrisico's die daar uit voortvloeien, eigenaarschap en accountability van big data en big data samenwerkingsvormen. Het literatuuroverzicht wordt afgesloten met een synthese van de bevindingen in de vorm van een raamwerk.

2.1 Big data en privacyrisico's

2.1.1 De eigenschappen van big data

Zowel wanneer big data oplossingen en diensten worden ingericht als wanneer datadiensten worden geleverd moet er vertrouwen zijn. Tijdens het inrichten van governance moet er bijvoorbeeld gezorgd worden dat partijen elkaars data gebruiken en dat de privacy van individuen beschermd blijft. Wanneer oplossingen zijn ingericht, geldt dat er duidelijkheid moet zijn over de oorsprong en kwaliteit van data en dat, indien er een privacyschending heeft plaatsgevonden, er voldoende waarborgen zijn. De eigenschappen van big data verhogen echter de kans op privacyschendingen. Big data staat voor het gebruik van datasets die te groot zijn om met reguliere IT toepassingen te verwerken. Big data wordt gekarakteriseerd door een groot volume, hoge snelheid en grote variatie in data. Het koppelen van datasets leidt weer tot nieuwe data. De gedachtegoed van big data gaat uit van datamaximalisatie: hoe groter de datasets en hoe meer datasets met elkaar gekoppeld worden, hoe meer de potentiele waarde (bijvoorbeeld nieuwe toepassingen). De big data gedachtegoed is echter niet zonder risico's. We behandelen de privacyrisico's van big data aan de hand van drie delen van de big data procesketen: het verzamelen, analyseren en toepassen van big data. Binnen elk van deze drie onderdelen spelen verschillende aspecten in het kader van vertrouwen.

2.1.2 Verzamelen

Bij het verzamelen van gegevens speelt vaak de vraag of het is toegestaan. Als het persoonsgegevens betreft, mogen deze dan verzameld worden? En welke randvoorwaarden gelden daarbij? In beginsel kunnen deze vragen vanuit het juridische kader omtrent gegevensbescherming beantwoord worden. Zo moet er een legitieme grondslag voor de verwerking zijn, de verwerking moet een specifiek doel hebben, er gelden informatieverplichtingen richting de betrokkenen en er dienen adequate organisatorische en technische maatregelen getroffen te zijn om de gegevens te beschermen. Hoewel deze kaders op het eerste gezicht helder lijken, ontstaan er juist in de context van big data toepassingen problemen. Het doel waarvoor gegevens verwerkt worden is vaak nog niet vooraf geheel duidelijk. De correlatie van gegevens en de uitkomsten van analyses kunnen vernieuwende inzichten opleveren die niet (te) voorzien waren. Dat is juist één van de grote beloften van big data toepassingen. Ook informatieverstrekking richting degenen wiens data verzameld worden kan soms lastig zijn. En technische en

organisatorische maatregelen die moeten voorkomen dat gegevens buiten de oorspronkelijke context terecht komen, verliezen mogelijk hun waarde wanneer het delen en combineren van gegevenssets een belangrijk aspect wordt om de economische waarde van big data te optimaliseren.

2.1.3 Analyseren

Binnen de fase van analyseren zijn tal van verwerkingen van de data mogelijk. Het uitgangspunt voor de verwerking kan een gericht vastgesteld doel zijn. Een vooraf vastgesteld doel voor data kan bijvoorbeeld zijn om verkeersstromen in kaart te brengen om uiteindelijk files te kunnen voorspellen en op individueel niveau reisadviezen aan burgers te geven. Echter, het toepassen van statistische programma's en algoritmes om verbanden te ontdekken in data om nieuwe inzichten te verwerven is ook een optie van big data. Een voorbeeld van een toepassing waar een inzicht is herkend op basis van enorme hoeveelheden data is Google Flu Trends, waarmee griepepidemieën over de hele wereld in kaart worden gebracht.[3] Ongeacht of het exacte doel vooraf wel of niet duidelijk is, wordt er in de analysefase naar gestreefd om nieuwe verbanden te vinden die vervolgens kunnen worden toegepast.[4] In de analysefase kunnen ook datasets gecombineerd worden. Daardoor kan informatie met elkaar in verband worden gebracht die eerst volledig los van elkaar stond. Met betrekking tot privacy zijn er in deze fase twee mogelijkheden. De combinatie van data kan, ook als begonnen wordt met geanonimiseerde gegevens, tot (her)identificatie leiden. Anderzijds is het ook mogelijk dat de enorme omvang van de datasets en de grote hoeveelheid records die zich in de set bevinden leidt tot een betere privacybescherming. De anonimiteit van individuen neemt immers toe naarmate er meerdere personen in de dataset zitten die aan eenzelfde profiel voldoen.[5]

2.1.4 Toepassing

Wanneer uitkomsten van analyses op big data worden toegepast kan eigenlijk pas duidelijk worden wat de daadwerkelijke impact van de gegevensverwerking is op individuele personen of op groepen. Een belangrijk deel van het vertrouwen hangt daarom samen met de toepassing, de gevolgen daarvan en de perceptie van het publiek over de toepassing. Het is belangrijk dat de uitkomsten, bijvoorbeeld aankoopvoorspellingen op basis van profielen, een gemiddelde vertegenwoordigen en vaak beïnvloed zijn door de zoekopdrachten die zijn uitgevoerd in de database. Daarmee zijn de uitkomsten niet altijd van toepassing op alle personen die aan een profiel voldoen en ook niet objectief. dana boyd en Kate Crawford geven deze subjectiviteit van big data aan in hun kritische essay over big data: "In reality, working with Big Data is still subjective, and what it quantifies does not necessarily have a closer claim on objective truth …"[6] Deze subjectiviteit is meteen één van de belangrijkste risico's van toepassingen van big data als basis voor bepaalde voorspellingen en beslissingen. Voor bedrijven zal namelijk vaak de hogere opbrengst door betere beslissingen of voorspellingen opwegen tegen een relatief beperkt aantal negatieve gevolgen op individueel niveau. Het verbod op

---

[3] http://www.google.org/flutrends/about/how.html.
[4] De privacy impact van Big Data. Considerati 2013, p. 6.
[5] L. Sweeney. *k*-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems,* 10 (5), 2002; 557-570. K-Anonymity betekent dat er voldoende overlap in personen in een databank is om te voorkomen dat een record aan één individu gekoppeld kan worden.
[6] danah boyd & Kate Crawford, Critical Questions for Big Data, INFO. COMM. & SOC'Y (MAY 2012), p. 6.

geautomatiseerde beslissingen zoals vastgelegd in Artikel 42 Wbp speelt dus een belangrijke rol in de toepassingsfase van big data.

2.1.5 Regelgeving ten opzichte van de fasen in de big data keten

Het lijkt erop dat regelgeving over gegevensbescherming vooral kaders biedt voor de eerste fase van big data; het verzamelen van data. Daarna ontstaat echter een afstand tussen de technologische praktijk en regelgeving, met name wanneer door bijvoorbeeld anonimisering de data geen persoonsgegevens meer zijn. Anonimisering is echter geen garantie voor de toekomst, in die zin dat combinaties van datasets of analyse van beschikbare data alsnog of wederom tot identificeerbaarheid kunnen leiden. Zo zijn er voorbeelden van geanonimiseerde datasets die met behulp van publiek beschikbare bronnen opnieuw gekoppeld zijn aan individuele personen. Zodra er één gegeven aan een identificeerbaar persoon is gekoppeld, leidt elke associatie van anonieme data met dat gegeven tot het opheffen van de anonimiteit van deze data.[7]

Daarnaast kunnen toepassingen van big data ook zonder heridentificeerbaarheid een behoorlijke impact hebben op individueel niveau, bijvoorbeeld omdat een individu in een bepaalde categorie wordt geplaatst of omdat een maatregel op algemeen niveau wordt toegepast. De grootste privacy-impact kan dus ontstaan in de toepassingsfase van big data, terwijl de juridische kaders omtrent gegevensbescherming daar momenteel niet of nauwelijks op aansluiten. Bovendien leiden de onvoorzienbaarheid van toepassingen en het mogelijk initieel werken met anonieme gegevens tot een gebrek aan invulling van alle waarborgen, zoals een rechtmatige grondslag en het vervullen van informatieverplichtingen. Het is in dergelijke gevallen immers niet mogelijk om een duidelijke doelomschrijving te formuleren op basis waarvan bijvoorbeeld toestemming voor de verwerking van persoonsgegevens verkregen kan worden.

Ook algemene beginselen van het beschermen van persoonsgegevens komen in het gedrang. Bedrijven zijn bijvoorbeeld verplicht om het verzamelen en verwerken van persoonsgegevens te beperken tot alleen dat wat noodzakelijk is voor het gerechtvaardigde doel van de verwerking. Bovendien moeten gegevens wanneer deze niet meer nodig zijn voor het primaire doel vernietigd worden. Het big data business model sluit dan ook niet aan bij het beginsel van dataminimalisatie. Het verzamelen van meer data voor een langere periode wordt namelijk door het big data business model aangemoedigd: de meerwaarde van big data zit juist in die onvoorziene secundaire toepassingen, de ‘kroonjuwelen’ van big data.[8]

2.2 Eigenaarschap en accountability van big data als aanvullende maatregelen

De huidige regelgeving sluit dus onvoldoende aan bij de praktijk van big data. Daarom is het interessant om te kijken waar aanvullende beleidsmaatregelen kunnen leiden tot waarborgen om alsnog de privacy van individuen te beschermen, zonder dat de kansen die big data toepassingen bieden volledig teniet worden gedaan. Die aanvullende maatregelen liggen op het gebied van accountability en eigenaarschap. Zowel vanuit het perspectief van privacy, als vanuit eigenaarschap

[7] A. Narayanan & V. Shmatikov, Robust De-anonymization of Large Sparse Datasets, 2008 IEEE Symposium on Security and Privacy, p. 119.
[8] O. Tene & J. Polonetsky. Big Data for All: Privacy and User Control in the Age of Analytics, p.22. Beschikbaar via SSRN: http://ssrn.com/abstract=2149364.

en accountability, worden er randvoorwaarden gesteld aan de inrichting van big data oplossingen en diensten. De voorwaarden uit de Wbp zijn hiervoor al genoemd.

2.2.1 Eigenaarschap

Eigenaarschap, ook wel zeggenschap, gaat over de rechten en verantwoordelijkheden die organisaties en individuen hebben ten aanzien van bepaalde datasets en het combineren daarvan. Bij eigenaarschap speelt dat degene die over de data beschikt, mogelijk een databankenrecht, auteursrecht, of bepaalde licenties heeft. De eigenaar heeft dus bepaalde aanspraken, maar ook bijbehorende verantwoordelijkheden. Een eigenaar heeft immers te zorgen dat er geen schade ontstaat voor anderen als gevolg van het gebruik van zijn eigendom. Wanneer een partij (toegang tot) een dataset verkrijgt van een andere partij valt dit binnen het kader van het verzamelen van data. Voor de verstrekkende partij is het echter een vorm van data toepassing. Daarnaast is het ook vanuit economisch perspectief belangrijk dat organisaties de data die ze hebben benutten om nieuwe waarde te creëren. Dat houdt ook in dat data gedeeld kunnen worden met anderen om combinaties mogelijk te maken. In deze gevallen dient er een balans te worden gevonden tussen de belangen van de eigenaar van de data (of in ieder geval degene die er iets mee wil) en de privacy van de personen wie de data betreft of van wie identificatie mogelijk wordt door het combineren van data. De mate waarin problemen optreden rond eigenaarschap en de rechtmatigheid van gegevensverwerkingen bij delen van datasets kan verschillen,[9] afhankelijk van het type samenwerking en de aanwezigheid van een eventuele hiërarchie.

2.2.2 Accountability

Accountability, dat aansluit op de verantwoordelijkheid van bedrijven zoals genoemd in de kabinetsvisie ePrivacy, is een centraal aspect om te verantwoorden hoe met data omgegaan wordt.[10] Bij accountability wordt vooral gekeken naar de wijze van verantwoorden van activiteiten, het verzamelen en gebruiken van gegevens en waarom een partij dat heeft gedaan. Accountability gaat in onze benadering dus over meer dan toerekenbaarheid van eventuele fouten of gebreken in de data. Ook aandacht voor eventuele gevolgen van datagebruik en het bieden van randvoorwaarden en remedies valt eronder. Voor accountability is het noodzakelijk dat er transparantie is over waar data vandaan komt en dat het mogelijk is om de oorsprong van data na te gaan, bijvoorbeeld via audits. Daarnaast speelt ook nog dat data verschillend van aard kan zijn. Zo is er data die bewust door individuen is verstrekt, is er metadata over klikgedrag en is er afgeleide informatie doordat datasets worden gecombineerd en profielen worden opgesteld. Ten slotte speelt ook het type toepassing een rol voor de eisen die aan de verwerking van persoonsgegevens worden gesteld. Al deze zaken zijn van invloed voor het inrichten van big data oplossingen, en deze verschillen in bepaalde situaties, zoals B2C en B2B.

2.3 Samenwerkingsvormen als controle

Hoe partijen samenwerken rond big data is essentieel voor de effectiviteit van het verzamelen, verwerken en toepassen van big data. Organisaties werken immers

[9] Powell, 1990
[10] Accountability is ook een belangrijk uitgangspunt in de voorgestelde Algemene Verordening Gegevensbescherming die op dit moment op EU niveau wordt vastgesteld.

samen om middelen efficiënter te gebruiken, om complementaire middelen te delen (bijvoorbeeld financiën, kennis, informatiesystemen, data, etc.) of om nieuwe diensten of producten te ontwikkelen en te exploiteren (innovatie). Met name middelen die schaars of kostbaar zijn worden door organisaties gedeeld in samenwerkingsverbanden. Big data en middelen om big data te verzamelen, analyseren en toe te passen zijn kostbaar. Daarom werken organisaties steeds meer samen om deze middelen efficiënter en effectiever te benutten zodat ze een competitief voordeel of meer maatschappelijke impact bereiken.

2.3.1 Typen samenwerkingsvormen en controle

Het ontstaan van organisatie, ook wel hiërarchie, als samenwerkingsvorm wordt gedreven door de een wens om transactiekosten tussen organisaties in de markt te verminderen. Echter, als de vermindering in transactiekosten niet meer opweegt tegen de organisatiekosten (bijvoorbeeld door bureaucratie), dan kan er weer worden gekozen voor de vrije markt i.p.v. hiërarchie. In de jaren negentig is, naast markt en hiërarchie, het organisatienetwerk als samenwerkingsvorm onder de aandacht van onderzoekers gekomen. Een netwerk bestaat uit drie of meer autonome organisaties die samenwerken om niet alleen individuele maar ook collectieve doelen te behalen (Provan & Kennis, 2008). Een netwerk is dus een vorm van collectieve actie: een sociale organisatie dat meer waarde creëert dan de som van de deelnemers en hun verbanden (O'Toole, 1997). Een goed netwerk biedt zowel publieke als private voordelen: organisaties leren van elkaar, er wordt efficiënter gebruik gemaakt van middelen, er is meer capaciteit om complexe problemen op te lossen, het biedt organisaties een competitief voordeel, en helpt om betere diensten te ontwikkelen voor klanten (Provan & Kenis, 2008). De coördinatie van samenwerkingsvormen bepaalt in belangrijke mate de uitkomsten van de samenwerking op organisatie-overstijgend niveau (Provan & Kenis, 2008). Deze coördinatie bestaat uit afgesproken *instituties* en *structuren* die moeten zorgen dat deelnemers betrokken blijven bij het collectieve doel, dat conflicten onderling worden beslecht, en dat de middelen van de samenwerking op efficiënte en effectieve manier worden gebruikt.

Een samenwerkingsvorm kan vervolgens op verschillende manieren worden gecoördineerd, variërend in mate van controle tussen de partijen:

- **Markt.** Bij markt als samenwerkingsvorm staat de autonomie van de deelnemende partijen centraal. De partijen hoeven elkaar niet volledig te vertrouwen: de samenwerking gaat via contracten en formele, dyadische transacties. Conflicten worden immers niet onderling beslecht, maar via de rechterlijke macht. Daardoor is, ondanks de afwezigheid van vertrouwen, de mate van controle hoog. Bij een markt als samenwerkingsvorm speelt de identiteit van deelnemers nauwelijks een rol.

- **Bazaar.** Reputatie en gemeenschapszin staan centraal in de bazaar als samenwerkingsvorm. Naar analogie van de Oosterse bazaar werken partijen onafhankelijk van elkaar aan een product of dienst. Via een open licentie zien partijen af van eigenaarschap, stimuleren ze de zo wijd mogelijke verspreiding van het product of dienst, en oefenen ze controle uit door transparantie en reputatie in de gemeenschap. Op deze manier is de bazaar als samenwerkingsvorm een alternatief voor markt: zonder formele contracten en zonder de behoefte aan vertrouwen werken partijen samen aan een dienst of

product. De identiteit van partijen is slechts gematigd van belang voor de gemeenschapszin en reputatie.

- **Hiërarchie.** Bij hiërarchie als samenwerkingsvorm staan de formele machtsverhoudingen tussen de deelnemende partijen centraal. Daardoor is er een hoge mate van controle door middel van sancties of beloningen. Deelnemers kunnen er voor kiezen om de besturing van het netwerk over te laten aan een of meerdere leidende organisaties of de oprichting van een paraplu organisatie. Bij het aanstellen van een parapluorganisatie om het netwerk te coördineren groeit het vertrouwen en consensus. De hiërarchische samenwerking wordt gekenmerkt door centraal bestuur (namelijk de dominante organisatie of de parapluorganisatie), een lage dichtheid van vertrouwen in het netwerk en een lage tot gemiddelde mate van consensus onder de deelnemers. Het plichtsmatige karakter van deze samenwerkingsvorm zorgt er voor dat de identiteiten van de deelnemende partijen van ondergeschikt belang zijn.

- **Netwerk.** In een lateraal organisatienetwerk coördineren deelnemers gezamenlijk de activiteiten, beslissingen, verdeling van de middelen en conflicten binnen de samenwerking. Onderling vertrouwen staat centraal: deelnemers gaan een sociaal contract met elkaar aan. Deze samenwerking wordt gekenmerkt door decentraal bestuur, een hoge dichtheid van vertrouwen in het netwerk en een hoge mate van consensus onder de deelnemers. De identiteit van de deelnemende partijen is wegens vertrouwen zeer van belang.

Tabel 1 geeft een overzicht van de eigenschappen van de samenwerkingsvormen.

Tabel 1 Eigenschappen van samenwerkingsvormen[11]

	Markt	**Bazaar**	**Hiërarchie**	**Netwerk**
Normatieve basis	Intellectueel eigendom	Open licentie	Arbeidsrelatie	Sociaal contract
Belang van identiteit van partners	Niet	Gematigd	Niet	Groot
Drijfveren	Competitie	Reputatie in de gemeenschap	Carrière	Vertrouwen
Redenen	Lage coördinatie-kosten en hoge flexibiliteit in deelname	Lage ontwikkelings-kosten en innovatie	Onderhande-lingspositie en differentiatie	Goedkope toegang tot middelen en gezamenlijk probleem oplossen

[11] Op basis van:
- Provan, K.G., en P. Kenis. Modes of network governance: Structure, management, and effectiveness. Journal of public administration research and theory 18.2 (2008): 229-252.,
- Lowndes, V., en C. Skelcher. The dynamics of multi□organizational partnerships: an analysis of changing modes of governance. Public administration 76.2 (1998): 313-333
- Demil, B., en X. Lecocq. Neither market nor hierarchy nor network: The emergence of bazaar governance. Organization studies 27.10 (2006): 1447-1466.

	Markt	**Bazaar**	**Hiërarchie**	**Netwerk**
Controle over de drijfveren	Hoog: via rechterlijke macht	Laag: via reputatie in community	Hoog: administratieve controle	Gemiddeld: reciprociteit en onderlinge controle
Flexibiliteit van samenwerking	Hoog	Hoog	Laag	Gemiddeld
Duur samenwerking	Eenmalig	Ongelimiteerd	Ongelimiteerd	Lange termijn
Toon van samenwerking	Achterdocht	Informeel, gericht op product	Formeel, bureaucratisch	Informeel, gericht op gezamenlijk belang
Verhouding van partners tot elkaar	Onafhankelijk	Gedeeltelijk afhankelijk	Afhankelijk	Onderling afhankelijk (reciprociteit)

2.3.2 Kiezen voor een big data samenwerkingsvorm

Welke samenwerkingsvorm wordt gekozen hangt af van factoren als netwerkgrootte, voorgeschiedenis, doel van de samenwerking, type data of informatiesysteem, openheid van het netwerk, netwerkhomofilie en fase van de samenwerking:

- **Netwerkgrootte:** hoe groter het netwerk, hoe belangrijker controle wordt over de deelnemende partijen (Provan & Kenis, 2008). Als een klein aantal deelnemers samenwerkt, dan kan vertrouwen als basis dienen voor de samenwerking. Naar mate het netwerk groeit, wordt het moeilijk om het gedrag van alle deelnemers te overzien en is meer contractuele of hiërarchische controle nodig.

- **Voorgeschiedenis van deelnemers:** eerdere samenwerking zorgt voor vertrouwen tussen (potentiële) deelnemende partijen. Een positieve voorgeschiedenis maakt het dus makkelijker voor deelnemers om voor een samenwerkingsvorm op basis van vertrouwen te kiezen.

- **Netwerkdoel** (bv. efficiëntie van middelen of innovatie): verschillende samenwerkingsvormen passen beter bij een bepaald netwerkdoel. Vertrouwen in het netwerk is essentieel voor innovatie, omdat concurrentiegevoelige materie wordt ontwikkeld bij de introductie van nieuwe producten of diensten. Het netwerk model (met name gedeelde besturing of een parapluorganisatie) en bazaar model passen bij een netwerk dat gericht is op innovatie. Aan de andere kant past een netwerk met meer controle (bv. hiërarchie of markt) goed bij samenwerking gericht op het delen van risico's en efficiëntie.

- **Type informatiesystemen en data:** de coördinatievorm kan gekozen worden doordat het aansluit bij het informatiesysteem dat gezamenlijk wordt ontwikkeld, geïmplementeerd en beheerd. Er is een verband tussen de structuur van het informatiesysteem en hoe de samenwerkingsvorm wordt ingericht. Bij gedeelde informatiesystemen met grote risico's op het gebied van concurrentie, privacy, eigendom, fraude, etc. sluiten samenwerkingsvormen met een hoge mate van controle het best aan.

- **Open vs. gesloten netwerken:** markt en bazaar zijn open samenwerkingsvormen, terwijl (in mindere mate) netwerk en hiërarchie gesloten samenwerkingsvormen zijn. Daarnaast kan een netwerk zo worden ingericht dat het open is voor nieuwe deelnemers of juist niet. In elk geval moet nagedacht worden over effectieve adoptie en legitimiteit van nieuwe deelnemers.

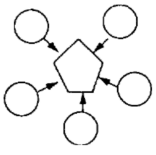
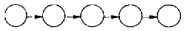
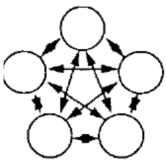
- **Netwerkhomofilie:** er kan spanning of juist overeenkomst tussen de individuele doelen van deelnemers zijn en de doelen van het netwerk. Een hogere mate van spanning vergt meer controle in de besturing van het netwerk. Legitimeren van de samenwerking naar de eigen organisatie, en de buitenwereld (bv. naar de media).

- **Fase van samenwerking**: de coördinatie van samenwerking is geen statisch gegeven. Gedurende de tijd veranderen de grootte van het netwerk, de doelen, etc. Daardoor zal bewust of onbewust de sturing veranderen met de structuur en richting van het netwerk.

2.4 Raamwerk: eigenaarschap en accountability in samenwerkingsvormen

Als organisaties samenwerken op het gebied van big data, dienen ze accountability en eigenaarschap onderling af te stemmen. De samenwerkingsvormen passen in meer of mindere mate bij de controle die organisaties willen houden op het verzamelen, analyseren en toepassen van de data. Tabel 2 geeft aan hoe het delen van big data verschilt in de verschillende samenwerkingsvormen.

Tabel 2 Eigenschappen van het delen van big data in samenwerkingsvormen[12]

Samenwerkingsvorm	**Markt**	**Bazaar**	**Hiërarchie**	**Netwerk**
Type informatie-systeem	Gebundeld 	Open source 	Keten 	Genetwerkt 
Typering van het delen van data	Verkopen en kopen van data per transactie	Vrijgeven en hergebruik van data (open data)	Op aanwijzing van centrale organisatie data leveren en gebruiken	Onderling delen van data

[12] Type informatiesysteem is gebaseerd op Kumar, K., en H.G. Van Dissel. Sustainable collaboration: managing conflict and cooperation in interorganizational systems. Mis Quarterly (1996): 279-300.

Samenwerkings vorm	**Markt**	**Bazaar**	**Hiërarchie**	**Netwerk**
Mate van controle over uiteindelijke toepassing van data	Laag wegens niet-duurzame relatie	Bewust laag	Hoog, wegens dominante partij	Gemiddeld, vanwege duurzame samenwerking en reciprociteit
Accountability over big data				
Verzamelen	Iedere partij afzonderlijk	Iedere partij afzonderlijk	Centrale (dominante of paraplu) organisatie	Deelnemers gezamenlijk
Analyseren	Iedere partij afzonderlijk	Iedere partij afzonderlijk, maar met terugkoppeling naar andere partijen	Centrale (dominante of paraplu) organisatie	Gezamenlijk afgestemd
Toepassen	Iedere partij afzonderlijk	Iedere partij afzonderlijk	Centrale (dominante of paraplu) organisatie	Deelnemers in gezamenlijke afstemming

Typering van het delen van data
In de markt als samenwerkingsvorm staat de dyadische transactie centraal: het kopen en verkopen van big data. Er hoeft daarom geen duurzame relatie tussen de partijen te zijn: als er naar verloop van tijd een betere transactie in de markt mogelijk is, dan zal deze de voorkeur krijgen. Bij de bazaar als samenwerkingsvorm staat het openstellen van data voor verdere hergebruik centraal, ook wel bekend als 'open data'.[13] De partijen vormen bij bazaar een gemeenschap rondom de data, die elkaar stimuleert om meer data vrij te geven en toepassingen te ontwikkelen. In een hiërarchische situatie wordt het delen van data opgedragen en gecontroleerd door een centrale organisatie. Als partijen in een laterale verhouding samenwerken (het netwerk als samenwerkingsvorm) staat het onderling delen van data centraal, waarbij partijen uitgaan van wederkerigheid. Zowel in een hiërarchie als in een netwerk hebben partijen een duurzame relatie met elkaar.

Eigenaarschap
Bij een netwerk of marktsamenwerkingsvorm blijft het eigenaarschap van de data bij de afzonderlijke partijen, waarbij middels licenties afspraken worden gemaakt over het delen van de data. Een voorbeeld is de exclusiviteit van de data. In een hiërarchische samenwerkingsvorm heeft de centrale organisatie eigenaarschap over de data. In de bazaar samenwerkingsvorm zien partijen via een open licentie juist af van eigenaarschap: het zoveel mogelijk verspreiden en toepassen van de data staat immers voorop. Een open licentie is echter niet mogelijk wanneer data persoonsgegevens bevat of (indirect) te herleiden is tot persoonsgegevens.

[13] Huijboom, N.M., en T.A. Van den Broek. "Open data: an international comparison of strategies." European journal of ePractice 12.1 (2011): 1-13.

Accountability
De accountability over het rechtmatig verzamelen, analyseren en toepassen van data ligt bij markt en bazaar bij de individuele deelnemers van de samenwerkingsvorm. Ongeacht wie de samenwerkende partij is, ieder moet zelf kunnen verantwoorden dat de data conform de Wbp wordt opgeslagen, verwerkt en toegepast. In het verlengde daarvan moet, in onze opvatting van accountability, iedere partij dan ook zelf zorgdragen voor zorgvuldige verwerking van de data, het inbouwen van maatregelen om schade voor individuen te voorkomen, en, indien toch schade mocht ontstaan, het bieden van remedies. Bij een hiërarchie ligt deze verantwoordelijkheid primair bij de centrale organisatie. De afhankelijkheid van organisaties tot de centrale organisatie is dus functioneel: zij moeten verantwoording afleggen aan de centrale organisatie. In een netwerk dient accountability tussen de partijen worden afgestemd. Aangezien deze samenwerkingsvorm wordt gekenmerkt door een hoge mate van vertrouwen, dienen partijen extra aandacht te besteden aan het gezamenlijk verantwoorden van het rechtmatige gebruik van data en het bieden van waarborgen.

Balans tussen controle en vrijheid in het toepassen van data
Vanuit het perspectief van de individuele organisatie is zoveel mogelijk controle nodig op de toepassing van data door derden om zowel privacyrisico's als afbreukrisico's uit te sluiten. Dat maakt samenwerkingsvormen met een hoge mate van afhankelijkheid en duurzame relaties wenselijk om controle te houden op het gedrag van de deelnemende partijen. Hoe gevoeliger de data, bijvoorbeeld medische informatie, hoe hoger de risico's bij privacyschendingen en des te wenselijker het is om meer controle te hebben over de toepassing van de data. Aan de andere kant maakt meer controle een gelijkwaardige samenwerking, met bijvoorbeeld innovatie als doel, lastiger. Vanuit het perspectief van de gebruiker is juist een mate van user empowerment nodig om privacy zoveel mogelijk te beschermen: als organisatie moet je transparantie geven aan gebruikers over hoe er met hun data wordt omgegaan. Kortom, om risico's en user empowerment t.a.v. privacy optimaal af te stemmen kan er gekozen worden voor een netwerk als samenwerkingsmodel, waarbij controle en user empowerment in balans zijn.

De concepten en relaties uit bovenstaande raamwerk zijn uitgewerkt in een vragenlijst voor een semigestructureerd interview (zie annex A voor het interviewprotocol).

3 Use cases

De use cases zijn geselecteerd op basis van theoretical sampling, waarbij er is gekozen voor praktijkvoorbeelden die lijken te passen bij elk van de vier typen samenwerkingsvorm: markt, bazar, hiërarchie en netwerk. Voor het markt type is gekeken naar Ahold personal marketing, voor de bazar naar Rotterdam open data, voor de hiërarchie naar Achmea Health Database en voor het netwerk type naar Eneco energie data. Elk van de use cases is gebaseerd op literatuuronderzoek en op tenminste één interview. Waar mogelijk is geprobeerd om met betrokkenen vanuit verschillende rollen te spreken. Vier aspecten per casus zijn uitgewerkt: big data samenwerking, eigenschappen van big data, privacyrisico's en eigenaarschap en accountability.

3.1 Ahold personal marketing[14]

3.1.1 Big data samenwerking

Ahold gebruikt (big) data vooral voor marketingdoeleinden binnen de eigen organisatie. Om al deze gegevens te koppelen, zijn alle retailers (de Albert Heijn vestigingen) aangesloten. Daarnaast werkt Ahold met één partij samen: Symphony EYC. Dit is een wereldwijde data analist die voor Ahold/Albert Heijn analyses uitvoert op de transactiegegevens om zo te bepalen welke producten klanten kopen en waarom. Ahold en Symphony EYC hebben dus een commerciële relatie: Ahold betaalt Symphony EYC om deze analyses uit te voeren. Op basis van de analyses die Symphony EYC uitvoert, doet Albert Heijn haar klanten gerichte aanbiedingen die als doel hebben klanten aan zich te binden en de omzet te laten stijgen.

3.1.2 Eigenschappen van big data

De belangrijkste datastroom zijn dus transactiegegevens van de aankopen die klanten doen in de Albert Heijn winkels. Deze gegevens van Albert Heijn zijn gekoppeld aan het 'Bonuskaart' loyaliteitsprogramma. Doel van het loyaliteitsprogramma is om klanten zo goed en gericht mogelijk aanbiedingen te doen van producten. Dit betekent dat producten worden aangeboden die klanten geneigd zijn om te kopen, maar ook dat het productadvies zo persoonlijk mogelijk wordt gemaakt. Daarnaast wordt ook het online aankoopgedrag (via Albert.nl) gekoppeld. Al deze transactiegegevens zijn zeer waardevol en analyses hierop geven al veel inzichten zonder dat er andere databronnen aan gekoppeld worden. Stapje voor stapje worden nu ook andere gegevensbronnen gekoppeld, zoals data uit marktonderzoek en demografische gegevens.

Omdat veel transactiedata gekoppeld zijn aan bonuskaartdata, hebben deze transacties een unieke identificatienummer. Op basis van deze identifier kunnen dan ook persoonlijke aanbiedingen worden gedaan. Er zijn op dit moment twee typen bonuskaart in omloop: online geactiveerde en niet geactiveerde kaarten. De geactiveerde kaarten zijn online geregistreerd door de gebruikers en zijn dan ook tot een persoon te herleiden. Deze personen ontvangen gericht aanbiedingen op basis van hun aankoopgedrag. In Nederland hebben 5 miljoen mensen dit gedaan. De niet geactiveerde kaarten zijn niet geregistreerd, maar hebben wel een uniek nummer. In tegenstelling tot de gegevens van de geactiveerde kaarten worden

[14] Gebaseerd op een telefonisch interview met Roland Tabor, Hoofd Personal Marketing bij Ahold

analyses op de gegevens van deze laatste groep worden niet op het niveau van het identificatienummer uitgevoerd, maar alleen op geaggregeerd niveau. Na de zomer wordt het ook mogelijk om bonuskaarten te koppelen, bijvoorbeeld de kaarten van verschillende gezinsleden, zodat er nog gerichter aanbiedingen gedaan kunnen worden.

De transactiegegevens van de geactiveerde bonuskaarten bevatten dus persoonsgegevens, omdat de transactie tot een unieke klant is terug te leiden. De data vanuit alle verschillende winkels wordt in grote databases centraal opgeslagen. Het koppelen en verzamelen van deze gegevens is behoorlijk complex. Zo moeten alle kassa's worden aangepast aan het verzamelen van data en is het lastig om te bepalen welke data allemaal over bananen gaan en hoe deze aan elkaar te koppelen. Ook moeten processen zo worden ingericht dat er elke week gerichte bonuskaart aanbiedingen worden gedaan. Zelfs het personeel moet hiervoor worden opgeleid. Zo mag een klant volgens de Nederlandse privacywetgeving niet actief worden benaderd om persoonsgegevens af te geven.

3.1.3 Privacyrisico's

Op basis van de transactiegegevens wordt profiling gedaan bij Symphony EYC. Dit gebeurt op basis van segmentatie van de Albert Heijn klanten in zes verschillende profielen. Maar er worden ook analyses gedaan om te bepalen welke producten het beste bij welke klant passen zodat deze aanbiedingen ook gericht aan bepaalde klanten gestuurd kunnen worden. Hiervoor worden de transactiegegevens van de geactiveerde kaarten op basis van een versleutelde identifier naar Symphony EYC gestuurd. Wanneer de analyses zijn gedaan worden deze opnieuw met deze versleutelde identifier terug gestuurd naar Ahold. Vlak voordat gepersonaliseerde e-mails met gerichte aanbiedingen worden verstuurd, worden deze weer gekoppeld aan de unieke identifier, waar ook een e-mailadres aan gekoppeld is. Alleen mensen met een gepast autorisatieniveau hebben toegang tot deze gegevens. Ahold zorgt verder voor naleving van de Wbp door het aanstellen van een Chief Privacy Operater, het doen van stevige interne audits en ook door regelmatig externe audits te laten doen.

De belangrijkste barrières zijn op dit moment vooral technisch van aard. Het inrichten van dergelijke processen en systemen vergt grote investeringen, niet alleen financieel, maar ook qua personeel. Zo is gegevensbeveiliging steeds belangrijker. Waar data eerst werden verwerkt in een 'closed customer loop', verandert dit nu in een continue proces van verwerken en verrijken van gegevens en het doen van analyses. Daarnaast zijn er veel legacy systemen die verandering lastiger maken, des temeer omdat deze systemen in een voortdurend doorgaande retailomgeving moeten werken. Privacy is vooral een barrière omdat het beperkend werkt voor het doen van gerichte aanbiedingen aan klanten. Soms is het dan ook niet mogelijk om een klant nog meer te helpen, terwijl Ahold/Albert Heijn dat wel zou willen. Een voorbeeld is dat alle transactiedata worden weggegooid wanneer iemand een bonuskaart verliest vanwege privacywetgeving. Sommige klanten zouden deze transactiehistorie echter graag behouden omdat ze graag gerichte aanbiedingen willen ontvangen.

3.1.4 Eigenaarschap en accountability

De transactiedata die worden verwerkt zijn en blijven eigendom van Ahold/Albert Heijn. Bij de registratie van de bonuskaart, moet de klant expliciet toestemming

verlenen voor de verwerking van gegevens. Hiermee blijft de klant aan het stuur. Mensen die hun kaart niet willen registreren, worden hiertoe niet gedwongen. De klant kan daarnaast altijd vragen om het gegevens te laten verwijderen. Data die van anderen wordt verkregen, zijn doorgaans marketing gerelateerde analyses die worden gedaan in opdracht van Ahold/Albert Heijn. Deze worden dan ook eigendom van de organisatie. Gegevens worden niet verkocht aan derden.

Het invoeren van de bonuskaart en het doen van gerichte aanbiedingen per e-mail is pas een eerste stap die Ahold/Albert Heijn is ingeslagen op het gebied van de personal marketing. Op de langere termijn is de verwachting dat er meer mogelijk zal zijn, bijvoorbeeld door op basis van sensordata realtime en geografische data te koppelen. Uiteindelijk wordt daarmee echt één-op-één contact mogelijk. Zo kan er gedacht worden aan het maken van geautomatiseerde persoonlijke boodschappenlijstjes en het klaarzetten van boodschappentassen zodat producten alleen nog afgehaald (of thuisgebracht) hoeven worden.

3.2 Rotterdam open data[15]

3.2.1 Big data samenwerking

Het open data portal van de gemeente Rotterdam (Rotterdam Open Data) is opgezet door kenniscentrum Creating 010, onderdeel van de Hogeschool Rotterdam. De Hogeschool meende dat datasets een rol konden spelen in het projectonderwijs van de Hogeschool en startte daarom, met subsidies, twee onderzoeksprojecten. Het eerste project ging over het verkennen van de mogelijkheden voor het openen van datasets van vier afdelingen binnen de gemeente: Stadsbeheer, Stadsontwikkeling, Stadsarchief en de Bibliotheek. Vooral met Stadsbeheer is er vervolgens een goede samenwerking ontstaan en die afdeling heeft veel datasets geopend. Het tweede onderzoek ging over het beschikbaar stellen van data via een open data portal, een data store. Dit is Rotterdam Open Data geworden. De voornaamste doelen van het open data portal zijn om de efficiëntie van de gemeente te verhogen, bijvoorbeeld doordat Wob-verzoeken niet meer ad hoc behandeld hoeven worden in de toekomst, en om innovatie binnen de gemeente te stimuleren, bijvoorbeeld doordat er beter beleid wordt gemaakt. Daarnaast hoopt de gemeente ook dat open data innovaties bij andere organisaties teweeg kan brengen, bijvoorbeeld door dat app ontwikkelaars nieuwe diensten creëren. Ten slotte heeft het publiceren van gemeentelijke data als doel om transparanter te worden.

Traditioneel werd de data die nu open is gemaakt hergebruiken en verkocht, bijvoorbeeld aan de RET, de Rotterdamse openbaar vervoer organisatie, en het Havenbedrijf. Dit is was altijd een inkomstenbron voor de gemeente, die met de komst van open data verdwijnt. Mogelijk kunnen gemeentelijke diensten nog wel verdienen aan het verkopen van bewerkte data. ‘Nieuwe’ afnemers van open data vragen de gemeente soms om specifieke datasets, maar dit is geen constante stroom aanvragen. Er wordt op dit moment samenwerkt met startups en app ontwikkelaars zoals 2CoolMonkeys en Sense-OS. 2CoolMonkeys heeft de

[15] Gebaseerd op interviews met Ferry de Groot, Programmamanager Open Data, gemeente Rotterdam; Judith Lemmens, docent Hogeschool Rotterdam, kenniscentrum Creating 010; Karin de Goederen, adviseur bij Stadsbeheer, gemeente Rotterdam; Jan-Peter Larsen, directeur Sense-OS en Reind van Olst, mede-oprichter van 2CoolMonkeys.

Bomenspotter app ontwikkeld, waarin informatie over alle 180.000 bomen staat die worden beheerd door de gemeente.

Met Sense-OS, een bedrijf dat diensten ontwikkelt op basis van sensordata, werkt de gemeente aan smart city toepassingen, op basis van 'crowd sensing'. Hierbij worden er interpretaties gedaan die zijn geaggregeerd uit invididuele sensoren. Het bekendste voorbeeld hiervan is verkeersinformatie. Die wordt traditioneel verzameld op basis van lussen die in de weg liggen, maar nu wordt met TomTom en via sensordata in mobieltjes duidelijk waar het verkeer vaststaat en is het mogelijk om real-time de wachttijden in kaart te brengen. Typische voorbeelden hiervan zijn het monitoren van waar bussen op dit moment rijden of het monitoren van gewassen in kassen. Voor open data is Sense-OS overigens nog op zoek naar een geschikt business model, want het is niet duidelijk wie er voor de hosting gaat betalen.

De samenwerking met startups is de belangrijkste manier van samenwerking rondom open data die de meeste resultaten heeft opgeleverd, zoals de bomenspotter app, maar ook een hooikoorts app die per straat de intensiteit van hooikoorts laat zien, afhankelijk van het type begroeiing. Voor apps is het belang dat ze hiermee hun werk en diensten aan commerciële partijen kunnen showcasen. Een andere partner van de gemeente is esri, een commercieel bedrijf dat GIS-systemen ontwikkelt en dat veel gebruik maakt van open data. En esri werkt ook weer samen met 2CoolMonkeys, waarbij 2CoolMonkeys op basis van esri's geografische data apps ontwikkeld voor de klanten van esri.

Daarnaast verkent de gemeente 'samenwerking' of eigenlijk co-creatie met burgers, zoals het 'sociaal maken' van de datasets. Dit wordt gedaan door terugmeldfaciliteiten in te bouwen in diensten op basis van open data, of door tweets over een bepaald object in de openbare ruimte te laten zien. Hierdoor kan een interactief (in plaats van aanbodgedreven) platform ontstaan rondom een community. Wat interessant is, is dat burgers vaak andere wensen hebben voor data en diensten rondom de data dan gemeentelijke diensten als Stadsbeheer. Op dit verschil speelt esri bijvoorbeeld in, door net iets andere interpretaties te maken op basis van de gemeentelijke data. Bijvoorbeeld door dat Stadsbeheer vooral de oppervlakte van fietspaden wil weten om te bepalen hoeveel materie er nodig is om het wegdek te vervangen, terwijl esri de lengte publiceert, waar fietsers meer in geïnteresseerd zijn. Zo worden dus op basis van dezelfde data diensten ontwikkeld met vergelijkbare, maar niet dezelfde informatie. De kaarten van esri zien er dan ook vaak anders uit dan die van Stadsbeheer. Samen met 2CoolMonkeys is esri dan ook aan het verkennen of er apps gemaakt kunnen worden om burgers te betrekken bij het 'schouwen' van de omgeving.

3.2.2 Eigenschappen van big data

Op dit moment is de data die wordt gepubliceerd vooral afkomstig van de gemeente, zoals de geografische data van de gemeentelijke dienst Stadsbeheer. Deze dienst beheert alle objecten in de openbare ruimte en heeft zeer veel kaartinformatie. Andere partijen zetten nu geen data in het portal. Zo is bijvoorbeeld de Hogeschool Rotterdam zelf helemaal niet bezig met het openen van haar eigen data. Op dit moment betaalt de gemeente het beheer van het portal. Doel van de gemeente is om het portal op de langere termijn zelfstandig te kunnen laten bestaan en zichzelf te laten financieren. Dit kan betekenen dat straks ook andere

(semi-)publieke organisaties, zoals scholen en ziekenhuizen, data in het portal zullen zetten. Om dit te realiseren richt het portal zich momenteel minder op het toevoegen van nieuwe datasets en meer op het ontwikkelen van business cases voor het portal rondom specifieke thema's. Zo komt er binnenkort een app met daarin alle beschikbare parkeerplekken in partkeergarages, zowel van de gemeente als van Qparks.

In het open data portal is vooral vraag naar geografische data en real-time data (zoals data over verkeersstromen). Vanuit Stadsbeheer zijn de belangrijkste databronnen de metingen die de dienst al 150 uitvoert in de stad en de omgeving: de 'basisinformatie' van de dienst. Deze is opgeslagen in een eigen beheerssysteem die alle metingen en objecten in een kaart laat zien en waarop geklikt kan worden om meer informatie over een object te krijgen. Het gaat bijvoorbeeld om bankjes, verkeersborden en bomen. Financiële data wordt (nog) niet gepubliceerd omdat dit wordt gezien als gevoelige informatie en persoonsgegevens ook niet. Afhankelijk van het type data wordt een veel gebruikt formaat gekozen. Voor database/excel is dit vaak SQL, voor geografische data .csv. Behalve dat open datasets op het portal gezet worden, wordt er metadata toegevoegd om de data gemakkelijker vindbaar en doorzoekbaar te maken.

3.2.3 Privacyrisico's

Omdat er geen persoonsgegevens worden gepubliceerd, zijn er in eerste instantie geen privacyrisico's. Wel kunnen er mogelijk risico's optreden wanneer data worden gecombineerd met andere datasets waardoor er persoonsgegevens zouden kunnen ontstaan. De BAG-gegevens zijn voor een partij als Albert Heijn mogelijk heel waardevol, omdat ze gecombineerd met de gegevens van de bonuskaart ineens hele interessante inzichten opleveren. Om de privacy te bewaken bij het laten zien van de beschikbare parkeerplekken in de stad, wordt de actuele informatie slechts één dag bewaard.

Stadsbeheer zou graag meer gebruik maken van open data voor haar eigen werkzaamheden, vooral bij het doen van analyses, maar loopt hierbij sterk tegen privacy aan. Zo zou de organisatie graag analyses willen doen naar de link tussen fysieke objecten en gevoelens van veiligheid, of naar waar hondenbezitters wonen zodat de buurt daar beter op kan worden ingericht of het opruimen van hondenpoep beter kan worden geregeld, maar deze informatie wordt door het gemeentelijke belastingkantoor niet gedeeld. Alleen op geaggregeerd niveau kan data worden gedeeld.

3.2.4 Eigenaarschap en accountability

De verkenning van de Hogeschool wees uit dat vooral 'objectieve' gegevens geschikt zijn om te openen, zoals de locatie van bankjes of bomen op de kaart. Zodra er een interpretatieslag van de gemeente overheen gaat, vind de gemeente het al lastiger om de data te delen. Bijvoorbeeld de interpretatie die de gemeente doet over of een boom ziek is. De gevolgen daarvan zouden kunnen zijn dat mensen zich actief gaan bemoeien met de inhoud van die kwalificatie en ze bijvoorbeeld gaan protesteren tegen het omzagen van een boom terwijl de gemeente vindt dat die ziek is. Twee datasets die opvallen omdat ze iets meer aan de 'interpretatie'-kant zitten zijn de datasets die de politie heeft ingebracht: de locatie van overvallen en fietsendiefstal gebaseerd op informatie over aangiftes bij de politie.

Het is voor het open data portal onmogelijk om van te voren te bepalen welke combinaties met de data allemaal tot persoonsgegevens gaan leiden. Wellicht zijn locatiegegevens hierop een uitzondering, omdat die al snel tot persoonsgegevens kunnen leiden. Daarom moet er wellicht goed nagedacht worden over het gebruik van sensordata, zoals GPS-gegevens via mobiele telefoons. De Hogeschool doet onderzoek naar de organisatie van de ontsluiting van data. Nu wordt open data via een stichting ontsloten in een open data store (Rotterdam Open Data), maar vragen over wie de eigenaar van de data is en wie het beheer van de data doet zijn nog niet goed beantwoord.

Ook Stadsbeheer zit met vragen over eigenaarschap, maar vooral vanuit het oogpunt van datakwaliteit. Zo heeft de dienst kwalitatief hoogwaardige informatie nodig voor haar beheerstaak, bijvoorbeeld voor de werkzaamheden van haar eigen personeel. Een voorbeeld is de iThor app die stadsmariniers gebruiken. Dit was een pilot die werd ontwikkeld door een externe organisatie. Nadat deze was afgelopen, zijn alle IT-systemen en de data in beheer genomen door Stadsbeheer omdat de data die werden verwerkt te privacygevoelig werden geacht om door een externe partij te laten beheren. Een ander voorbeeld hiervan is het beheren van gegevens over leidingen in de stad (en vroeger ook in het havengebied) die mogelijke een veiligheidsrisico kunnen vormen. Dezen kunnen vanwege dit veiligheidsrisico niet geopend worden. Stadsbeheer heeft daarnaast hele hoge normen voor het in kaart brengen van deze leidingen (zoals chloorleidingen) en voert daarom eigen metingen uit, om te zorgen dat deze data zeer nauwkeurig zijn. Om dezelfde reden is het verzamelen van data van burgers interessant, maar ook lastig – zeker wanneer het om dit soort gegevens gaat waarbij de kwaliteit hoog moet zijn.

3.3 Achmea Health Database[16]

3.3.1 Big data samenwerking

Achmea heeft twee-en-een-half jaar geleden een *Kenniscentrum* opgericht dat onderzoek uitvoert naar nieuwe producten, mogelijkheden tot verbetering van het zorgproces, verhoging van de kwaliteit van de zorg, nieuwe zorgmethoden, etc. Het Kenniscentrum is een stafafdeling, direct onder de verantwoordelijkheid van de voorzitter van de divisie Zorg en Gezondheid. Daarnaast heeft het bedrijf de *Achmea Health Database* (AHD), voortgekomen uit de vroegere AGIS Health Database, bevat de gegevens van ongeveer 4.7 miljoen patiënten en valt formeel onder het Kenniscentrum. De AHD is een epidemiologische databron, waarin alle gegevens van het zorggebruik van verzekerden zijn verzameld; kort gezegd alle zorgconsumptie waarvoor betaald wordt. De data wordt dan ook primair opgeslagen en gebruikt voor declaratie en betalingsdoeleinden. Daarnaast kan voor zuiver wetenschappelijke doeleinden een beslag gedaan worden op de gegevens.

Per jaar wordt er ongeveer 20 keer een beroep gedaan op beschikbaarstelling van de gegevens uit de Achmea Health Database. Overwegend zijn dit verzoeken van onderzoekers aan onderzoeksinstituten (universiteiten, TNO), soms ook in het kader van projecten binnen het ZonMW programma, een heel enkele keer door de industrie (bv. de farmaceutische industrie). Alle aanvragen worden beoordeeld door een commissie die kijkt naar wetenschappelijke en maatschappelijke relevantie, adequaatheid van de vraagstelling, noodzaak tot beschikbaarheid van gegevens,

[16] Deze use case is gebaseerd op een interview met Barry Egberts, senior manager en hoofd van het Kenniscentrum van Achmea en desk research.

onderzoekaanpak en dergelijke. In deze commissie zitten ook artsen die de medische relevantie kunnen beoordelen. Voordat artikelen gepubliceerd mogen worden, moeten ze bovendien worden voorgelegd aan de beoordelingscommissie. Dan wordt gekeken naar de correctheid van de aanpak en de bevindingen. Dit wordt gedaan om te voorkomen dat Achmea nadelen ondervindt van slecht uitgevoerd onderzoek. In het geval dat bedrijven een verzoek doen op beschikbaarstelling van data wordt zorgvuldig gekeken naar de maatschappelijke relevantie achter de onderzoeksvraag zit en het niet alleen om economisch gewin voor het bedrijf in kwestie gaat.

Op de website van de Achmea Health Database is een beoordelingsformulier te vinden met de voorwaarden waaronder de data wordt gedeeld met wetenschappers. Deze criteria betreffen alleen de criteria voor het delen van informatie voor wetenschappelijk onderzoek. De criteria zijn onder te verdelen in de volgende onderwerpen:

- Opzet en vraagstelling onderzoek
- Kwaliteit en haalbaarheid
- Maatschappelijke relevantie
- Ethische en privacy aspecten
- Implementeerbaarheid
- Data-technische uitvoering

De beoordeling vindt plaats in de onderzoekscommissie, bestaande uit vertegenwoordigers van Achmea en universiteiten. Bij de beoordeling van de aanmelding worden de genoemde criteria als geheel afgewogen.

3.3.2 Eigenschappen van big data

Achmea maakt onderscheid tussen het interne gebruik van zorgdata en het externe gebruik. Ook intern zijn er allerhande procedures die moeten bijdragen aan een goede omgang met persoonlijke zorgdata. Wie heeft toegang tot welke data en hoe wordt dit geborgd? Achmea voert momenteel een audit uit om zeker te zijn dat procedures correct zijn en ook naleefbaar zijn.

De gegevens die Achmea beheert zijn in principe beschikbaar voor extern wetenschappelijk onderzoek. Commerciële partijen doen slechts mondjesmaat een beroep op deze data. Blijkbaar werpt de wetenschappelijke toetsing die Achmea hanteert een drempel op die niet makkelijk genomen wordt. Het interne gebruik van de gegevens wordt gemonitord en daar zijn strakke procedures over afgesproken: wie heeft toegang tot welke gegevens op grond van welke overwegingen?
Het zorggebruik omvat alle intra- en extramurale medische en paramedische zorg, inclusief psychiatrie. De geboden zorg is standaard gecodeerd, zoals ATC voor medicatie en DBC voor ziekenhuisbehandeling, met een groot aantal detailgegevens en de bijbehorende kosten. De gegevens worden geautomatiseerd en continu aangeleverd door de zorginstellingen en worden gecontroleerd op juistheid, alvorens toe te voegen aan de database. Achmea voert een zorgvuldig privacy en veiligheidsbeleid uit voor het beschikbaar stellen van data.

De Achmea Health Database is primair opgebouwd voor het betalingsverkeer van zorgverrichtingen voor verzekerden. De controle op de juistheid van de gegevens is daarom bijzonder hoog. Achmea stelt vanuit haar rol van maatschappelijke verantwoordelijkheid de (anonieme) gegevens beschikbaar voor wetenschappelijk

onderzoek. Daarmee is het secundaire doel het vergroten van de kennis in de gezondheidszorg en het bevorderen van innovatie en doelmatigheid in de zorg. Het onderzoek met AHD gegevens dient bij voorkeur uit te monden in publicatie in erkende (internationale) wetenschappelijke tijdschriften. Onderzoekers kunnen uit de Achmea Health Database anonieme data verkrijgen door het indienen van een aanvraagformulier.

Er worden uitsluitend geanonimiseerde of (via een Trusted Third Party) gepseudonimiseerde gegevens geleverd. Vaak is dit project Mondriaan, een onafhankelijke, dienstverlenende non-profit organisatie die zich richt op het faciliteren van wetenschappelijk-medisch onderzoek. Mondriaan wil een geavanceerde infrastructuur van zorggegevens bieden voor beter én meer onderzoek zoals (farmaco-) epidemiologisch en economisch onderzoek. Deze infrastructuur moet breed toegankelijk zijn voor alle wetenschappelijk onderzoek. Wanneer er een “klant” komt die op zoek is naar een dataset, kopen zij een licentie voor het product van project Mondriaan. Dit geeft ze nog geen toegang tot de gegevens zelf. Daarvoor wordt contact opgenomen met de eigenaars van de bron, zoals de Achmea Health Database.

Van belang voor de honorering van een aanvraag bij Achmea is dat het om een wetenschappelijk of statistisch onderzoeksdoel gaat, waarvoor de zorgverzekeraar (conform de Gedragscode Zorgverzekeraars van ZN) de gegevens van haar verzekerden beschikbaar mag stellen. Daarnaast is in de polisvoorwaarden tevens opgenomen dat de data voor statistisch onderzoek gebruikt mag worden.
De gegevens worden geanonimiseerd door de TTP zodat geen herleidbaarheid van gegevens op kan treden. Ook is het mogelijk eigen datasets via een Third Trusted Party met een pseudonimisatie procedure te koppelen aan de zorggegevens van Achmea. In het bestand zijn naast de gegevens van zorggebruik ook kenmerken van verzekerden en zorgverleners opgenomen, waardoor gedetailleerde analyses goed mogelijk zijn. De gegevens zijn verzameld over een periode van meer dan 12 jaar.

3.3.3 Privacyrisico’s

Met betrekking tot de privacydiscussie onderstreept Achmea het belang van een goede communicatie, transparantie en een proactieve houding naar buiten. Op het moment dat je in het defensief wordt gedrongen is er vaak een probleem. Het is belangrijk om goed uit te leggen waar gegevens voor gebruikt worden. Een privacyschending is een groot risico voor het imago van een onderneming. Patiënten geven met het accepteren van de polisvoorwaarden aan dat ze akkoord gaan met het benutten van hun gegevens voor wetenschappelijke doeleinden. Daarbij is het noodzakelijk voor de verzekeraar om deze data te verzamelen vanuit wettelijk verplichtingen waaronder de verantwoording, risicoverevening etc. Achmea onderkent dat het accepteren van de polisvoorwaarden niet gezien kan worden als een vorm van informed en explicit consent. Het individueel afstemmen van deze toestemming maakt de werkbaarheid van de Achmea Health Database echter onhaalbaar, omdat de kosten die daar mee gemoeid zijn, de opbrengsten sterk overschrijden. Over de manier van toestemming verlenen is al eens ophef ontstaan in de Telegraaf, toen bekend werd dat gegevens van verzekerden gebruikt konden worden voor wetenschappelijk onderzoek. Dit staat echter in de voorwaarden en naar aanleiding van het publiceren van dit nieuwsbericht heeft Achmea in hun aanpak veel bijval gekregen van experts op privacygebied. De experts hebben

betoogd dat Achmea voldoet aan de Wet bescherming persoonsgegevens (Wbp) en verdere toestemming daarom niet nodig en bovendien niet haalbaar is.

Zoals hierboven genoemd, voldoet Achmea wat betreft het waarborgen van privacy aan de Wbp. Het risico dat het delen van gegevens en deze laten bewerken door onderzoekers, de data herleidbaar maken op persoonsniveau is hierbij een onvermijdelijk risico. Dit kan altijd gebeuren in het proces van analyseren en toepassen van data, maar is bij wetenschappelijk onderzoek nooit een doel op zich. Strakkere privacyregelgeving zal hierbij niet helpen, omdat deze issues vaak niet zwart-wit zijn. Aangezien Achmea werkt met wetenschappelijk onderzoekers gaan zij hierin uit van de integriteit van de onderzoekers. Ze hebben hier nooit conflicten of schandalen gehad.

Met ZorgTTP zijn afspraken gemaakt over het pseudonimiseren en het beheer van de gegevens, wanneer deze data gedeeld wordt met onderzoekers. Deze afspraken betreffen de gehele keten van aanlevering, opslag en doorlevering van zorgdata aan derden. Achmea levert ook gegevens aan via Mondriaan en maakt daarin gebruik van de TTP waar Mondriaan over beschikt.

3.3.4 Eigenaarschap en accountability

In het geval van zorggegevens is eigenaarschap een discutabel begrip. Zorggegevens kunnen namelijk eigenaar zijn van de verzekerde, de zorgverlener of de zorgverzekeraar. Dit is een discussie die momenteel erg actueel is, aangezien de verzekerde steeds meer centraal wordt gesteld in het zorgproces, terwijl eerder de zorgprofessional centraal stond. Dit heeft ook implicaties voor medische data – steeds meer wordt dit gezien als eigendom van de verzekerde. Het feit dat eigenaarschap voor zorggegevens ambigue is, maakt het lastig om dit begrip verder in te richten dat dat op het moment het geval is.

Binnen de AHD is eigenaarschap niet verder ingericht dan in de securityvoorwaarden staat beschreven. Aan deze securityvoorwaarden moet de partij aan wie de data verleend wordt, voldoen. Dit houdt in ieder geval in dat gegevens na vijf jaar vernietigd dienen te worden door degenen aan wie de gegevens beschikbaar gesteld worden. Er is echter nog geen toezicht of dit daadwerkelijk gebeurd, omdat het niet te controleren lijkt. Het is verder onduidelijk of de personen aan wie de data verstrekt worden, daadwerkelijk eigenaar worden van de gegevens, of dat Achmea dit blijft.

Accountability is een begrip dat binnen de Achmea Health database niet is vastgelegd. Het is diffuus wie er verantwoordelijk is bij eventuele schandalen of misbruik van de data. Het is nog nooit voorgekomen dat er daadwerkelijk schandalen plaats hebben gevonden met data vanuit de Achmea Health database, maar indien dit wel het geval is, zal per situatie bekeken moeten worden wie hiervoor aansprakelijk is. Achmea is van mening dat dit in eerste instantie de partij zou moeten zijn die gebruik maakt van de data, omdat deze partij de data bewerkt. Op deze manier kunnen persoonsgegevens ontstaan, welke schade zouden kunnen opleveren voor groepen of individuen. Hierover is echter niets vastgelegd en er wordt uitgegaan van de integriteit van de wetenschappers die gebruik maken van de data.

Voor Achmea is het wel mogelijk om achteraf geen toestemming te verlenen voor de publicatie van een onderzoek. Publicaties gaan eerst langs Achmea voordat deze ingediend worden bij een tijdschrift. Achmea kan afzien van publicatie wanneer zij de resultaten van het onderzoek niet ondersteunen. In de praktijk komt dit echter amper voor.

Een bezwaar voor het verder inrichten van zowel eigenaarschap als accountability is de werkbaarheid van het proces van delen van data voor wetenschappelijke doeleinden. De Achmea Health Database is een initiatief dat voor Achmea geen economische waarde oplevert, maar wel een grote bijdrage kan leveren aan wetenschappelijk onderzoek. Achmea heeft echter niet de middelen om de regelementen strikter te maken of het toezicht aan te scherpen. Het verder inrichten van eigenaarschap en accountability brengt mogelijk meer rompslomp met zich mee, wat het voor Achmea Health Database lastig maakt om hanteerbaar te blijven. Een ander bezwaar is dat het verder inrichten van privacy, eigenaarschap en accountability niet altijd mogelijk is. Patiënten hebben recht op bescherming van hun privacy en sommige zaken kunnen daardoor niet verder ingericht worden. Een voorbeeld hiervan is de ophef die rondom een recente uitzending van Zembla rondom de bescherming van medische gegevens. Hierin kwam naar voren dat het wettelijk mogelijk is voor patiënten om te eisen dat hun aandoening(en) niet vermeld worden op de nota. Voor Achmea is dit vervolgens wel weer noodzakelijk voor de kostenverrekening en controles op rechtmatigheid van de declaratie. De zorgverzekeraar trekt hierbij vaak aan het kortste eind en zal zich moeten voegen naar privacy van de patiënt. Hierdoor wordt het verder inrichten van eigenaarschap, accountability en privacy vaak belemmerd.

3.4 Energie data[17]

3.4.1 Big data samenwerking

Energieleveranciers en netwerkbeheerders werken anno 2014 nog weinig samen om energie gerelateerde gegevens onderling of met andere partijen te delen. Zo geeft Eneco aan dat het nog geen data deelt met derden. Voorheen werkten energieleveranciers en netwerkbeheerders samen met de publieke sector om bijvoorbeeld energieverbruik in Nederland op een gedetailleerd niveau in beeld te brengen of de politie te ondersteunen bij het opsporen van wietplantages.

Data speelt echter een steeds belangrijkere rol in de energiemarkt: data over energievraag en aanbod zorgt bijvoorbeeld voor nieuwe toepassingen op het gebied van energie-efficiënte. Inmiddels zijn er plannen om meer data met elkaar te delen, bijvoorbeeld via een open energiedata platform. Het doel van het delen van energie data is om een bijdrage te leveren aan de energie transitie in Nederland. Door de Nederlandse overheid als mede-eigenaar van energiepartijen is er sprake van publieke data: de verzamelde energie relateerde gegevens zijn immers medegefinancierd door publieke middelen. Het toegankelijk maken en koppelen van deze data kan leiden tot innovatie in energie gerelateerde diensten en een hogere transparantie in energiegegevens voor consumenten. Energieleveranciers en netwerkbeheerders hebben innovatieafdelingen die een sterke drijfveer vormen voor innoveren op basis van energiedata.

[17] Deze use case is gebaseerd op interviews met Willem van den Bosch (TNO) en Thomas de Groen (Eneco)

De redenen waarom er tot nu toe beperkt wordt samengewerkt tussen energiepartijen zijn:

- Veel data, bijvoorbeeld van slimme meters, is nog in intern beheer.
- Het ontbreken van aantrekkelijke business cases voor het (openlijk) delen van energiedata: de middelen die nodig zijn om data te delen overstijgen vaak de opbrengsten. De opbrengsten die Big Data belooft worden in de praktijk nog niet gehaald.
- De risico's die verbonden zijn aan het aanbieden van mogelijk privacygevoelige data.
- De complexiteit om met meerdere grote partijen samen te werken, bijvoorbeeld de snelheid waarmee beslissingen worden genomen.
- Onvoldoende bewustwording, wiskundige kennis, ervaring en skills om waarde te halen uit Big Energiedata. Waardevolle analyses vergen een unieke combinatie van ICT, wiskundige en domeinkennis.

Om het delen van energiedata vorm te geven hebben TNO, Enexis en KPN het project Toegankelijke Energie Informatie (TEI) gestart in 2014. Het TEI project richt zich niet alleen op het vrijgegeven van energie gerelateerde data (bijvoorbeeld energieverbruik op postcodeniveau), maar wil een 'level playing field' op het gebied van energiedata creëren door samen met andere netwerkbeheerders een data platform op te richten. Deelnemers (zogenaamde Joint Innovation Partners) werken in een organisatienetwerk met elkaar samen, waarbij KPN als centrale partij de ICT infrastructuur en beheer van het platform verzorgt. Het is de bedoeling dat het een open netwerk wordt: het platform staat open voor nieuwe deelnemers, zowel aanbieders als gebruikers van energiedata. De doelgroep voor het platform zijn netwerkbeheerders, dienstontwikkelaars (bijvoorbeeld mobiele applicatie ontwikkelaars), energieleveranciers en overige gebruikers. De doelgroep voor het leveren en gebruiken van de data is bewust niet beperkt tot energiepartijen om trans sectorale innovatie te bevorderen.

3.4.2 Eigenschappen van big data

De komende jaren zal energiedata exploderen door smart grids, slimme meters of ketels, social media en de opkomst van decentrale energievoorzieningen. Eneco heeft op dit moment 30.000 slimme thermostaten ("TOON") in de markt, maar de ambitie is om de slimme meter bij alle 2,2 miljoen aansluitingen te installeren. Deze 'Big Data' wordt gezien als een essentiële drijfveer voor innovatie in de energiemarkt. Slimme ketels kunnen bijvoorbeeld uit zichzelf aangeven dat ze toe zijn aan onderhoud zodat ze niet volledig vervangen moeten worden. Het platform van het TEI project zal data bevatten van netwerkbeheerders (bijvoorbeeld facturering, levering van energie, etc.), slimme meter data (op een geaggregeerd niveau), telecomdata (bijvoorbeeld storingen in regio's) en overige open data, zoals geografische data over leidingen onder de grond.

Binnen Eneco staat nu nog het verzamelen en verwerken van data uit operationele informatiesystemen (CRM of logistiek) centraal. Er zijn op dit moment pilots om te experimenteren met de analyse van slimme meter data, bijvoorbeeld om voorspellende waarde uit de data te halen: predictive analytics. Veel data is op dit moment nog niet toegankelijk en vrij te verkrijgen. Het project TEI zal zich er juist op richten om data te delen met derde partijen. Energiedata verschilt in hoe openlijk het gedeeld kan worden en de kosten die voor de data worden gevraagd. Daarom

hanteert het TEI project een gelaagd model voor de openheid van data. Niet alle data binnen het TEI is dus open data.

Inmiddels heeft het TEI project 15 use cases voor het platform bedacht, waarvan 6 uiteindelijk in het project zullen worden uitgewerkt. De use cases variëren van het identificeren van energiestoringen tot het optimaal afstemmen van energievraag en aanbod. In dit stadium is de voorzienbaarheid van de diensten die op basis van de data ontwikkeld zullen worden nog hoog. Het koppelen van meerdere datasets is op dit moment nog in een experimentele fase. De nadruk op de verzameling en analyse van energiedata ligt op waarde halen uit huidige datasets in plaats van datamaximalisatie (zoveel mogelijk data verzamelen en koppelen om tot inzicht en innovaties te komen). Als er echter een sterke toename komt in de hoeveelheid energiedata (zoals hierboven beschreven) en de middelen om snel data te koppelen en analyseren, dan kan de aandacht wel verschuiven naar datamaximalisatie.

3.4.3 Privacyrisico's

De privacy gevoeligheid van energie data is afhankelijk van het type en aggregatieniveau van de data. Zo is energieverbruik op regioniveau zeer lastig te herleiden tot persoonsniveau, maar is slimme meter data beschikbaar tot op het niveau van 6 huishoudens. Met TOON als slimme thermostaat heeft Eneco zelfs energiedata op het niveau van het huishouden. Het TEI platform zal op termijn data bevatten van de slimme meter, en daarom zijn er privacyrisico's aanwezig. Zowel het TEI project als Eneco zijn zich bewust van de privacyrisico's van het delen van energiedata, en geven aan dat het voldoen aan de eisen van de Wet Bescherming Persoonsgegevens een essentiële voorwaarde voor innovaties op basis van data is. Er zijn immers strenge regels vanuit de overheid voor het verzamelen, verwerken en beheren van slimme meter data door marktpartijen. Netbeheerders mogen slimme meter data alleen zonder toestemming van de consument uitlezen voor de jaarnota, tweemaandelijkse rekeningoverzichten, bij een verhuizing of als een uitlezing voor beheer van het energienet. Het beheer en verwerking van de slimme meter data wordt gemonitord door het College Bescherming Persoonsgegevens (CBP) en de Autoriteit Consument en Markt (ACM). De respondenten geven aan dat niet alleen het wettelijke kader rond privacy van belang is, maar ook de perceptie van het publiek speelt een belangrijke rol in wat wenselijk is met slimme meter of slimme thermostaat data. Een recent voorbeeld uit de financiële sector is de geschokte publieke reactie op het voorstel van ING om met toestemming van de klant van klanten te verkopen aan derden. Een van de respondenten verwoordt het innovatie-privacy dilemma als volgt: "Je wilt niet te laat zijn, maar ook niet te vroeg en jezelf in de vingers snijden omdat consumenten nog niet klaar zijn voor de meer geavanceerde diensten en dit qua PR misschien zelfs wel verkeerd uit kan pakken."

Een zorg voor de toekomst is de mogelijkheid tot heridentificatie wanneer meerdere datasets worden gekoppeld. Het TEI platform zal bijvoorbeeld niet alleen energiedata bevatten, maar ook open data van andere bronnen zoals geografische informatie. Kortom, wanneer de hoeveelheid 'big energie data' toeneemt en gedeeld wordt kunnen de privacyrisico's toenemen.

3.4.4 Eigenaarschap en accountability

De respondenten gaven aan dat eigenaarschap van energiedata een punt van discussie is. Deze discussie speelt zowel tussen organisaties, als tussen organisaties en consumenten. In het TEI project vormen de deelnemers een lateraal netwerk en zijn samen eigenaar van het platform. De individuele partijen zien zich zelf als eigenaar van de datasets die ze zelf inbrengen. Vervolgens zullen middels contracten, data agreements, licenties worden verleend aan de gebruikers van de data. Dit zijn geen standaard contracten. Het eigenaarschap van nieuwe data dat ontstaat door het combineren van datasets, zogenaamde linked data, is echter nog onduidelijk. Bij een mogelijk conflict tussen deelnemers binnen het TEI project , bijvoorbeeld een contractbreuk, zullen geen directe juridische consequentie volgen. Als deelnemer aan het netwerk is het namelijk belangrijk om de onderlinge relaties niet te beschadigen of onderlinge conflicten in de publiciteit te brengen. Daarom zullen bij evt. conflicten partijen onderling, op een ad-hoc basis worden opgelost.

Op dit moment ligt de nadruk in het TEI project op de technische en commerciële haalbaarheid van het platform. Eigenaarschap heeft daarom nu geen prioriteit en zal pas een rol spelen als het platform wordt opgericht. Eneco deelt nog geen gegevens met derde partijen. De discussie welke organisatie eigenaar van de data is speelt bij Eneco voorlopig dus geen rol. Wel speelt bij zowel Eneco als het TEI project de vraag in hoeverre de energieconsument eigenaar is van de data die over hem of haar is verzameld. Zowel in het TEI project als bij Eneco wordt hier over nagedacht.

De deelnemers van het TEI project zullen op termijn accountability van de data moeten vormgeven. Als het energiedata platform van het TEI project van start gaat dan zal er een aparte beheersorganisatie (een *network administration organization*) worden opgericht. Alle deelnemers in het netwerk, inclusief KPN, zullen zitting nemen in deze beheersorganisatie om de belangen van hun organisatie te vertegenwoordigen. Niet alleen de beheersorganisatie zal toezien en toetsen of de data op juiste wijze wordt verzameld, verwerkt en toegepast. Het platform zal gecertificeerd worden en gecontroleerd worden door een externe partij. Eneco deelt nog geen data met externe partijen, en dus hoeft op netwerkniveau niet na te denken over het inrichten van data accountability.

Zowel het TEI project als Eneco geven aan over accountability richting de consument, bijvoorbeeld door meer user empowerment, na te denken. Eneco wil de consument meer informeren over de data die ze verzamelen, bewerken en toepassen. De vraag is alleen hoe deze transparantie op een voor de consument waardevolle manier kan worden ingericht. Op welke informatie zit de consument te wachten en hoe kan je deze informatie het beste communiceren? Deze discussie speelt in mindere mate bij data op geaggregeerd niveau. De respondent van het TEI project vraagt zich af of het platform de juiste partij is om informatie te verschaffen aan de consument, of dat deze verantwoordelijkheid dient te liggen bij de gebruikers van de data. Het TEI platform zal waarschijnlijk transparantie geven aan tussenpartijen die wel direct contact hebben met de eindgebruiker. Deze tussenpartijen kunnen in het geval van particuliere diensten (bijvoorbeeld gaslicht.com) vervolgens aan de eindgebruiker geven. Mogelijk kunnen authenticatiesystemen, zoals DigiD, worden gebruikt voor het geven van toestemming en het inzien van persoonlijk energiegegevens.

4 Cross-case analyse

De cross-case analyse richt zich op elk van de vier aspecten die per casus zijn uitgewerkt: big data samenwerking, eigenschappen van big data, privacyrisico's en eigenaarschap en accountability.

4.1 Big data samenwerking

In de praktijk zijn verschillende typen samenwerkingsvormen zichtbaar. Tabel 3 geeft een overzicht van welke samenwerkingsvormen worden gebruikt in elke case. Zowel Ahold personal marketing als Achmea Health Database hebben een hiërarchisch samenwerkingsmodel: beide partijen zijn de dominante en beslissende partij in de samenwerking en houden een sterke mate van controle over de data. De gemeente Rotterdam staat centraal in de Rotterdam open data casus. Echter, in plaats van op te treden als een dominante partij, ondersteunt de gemeente juist een bazaar model waarin de community van datagebruikers centraal staat. De energiepartijen in de energie data case hebben een (hiërarchisch) netwerk als samenwerkingsvorm, waarin een *Network Administration Organization* op termijn verantwoordelijk is voor de coördinatie van het dataplatform. In geval van een conflict of onduidelijkheid over eigenaarschap wordt in deze casus ook binnen het netwerk ad-hoc naar een oplossing gezocht om schade aan de samenwerking te voorkomen. In geen enkele case is er sprake van een marktmodel waarin op een open en commerciële wijze data wordt gedeeld. Een reden voor het ontbreken van een marktmodel kan zijn dat alle voorbeelden aangaven dat er nog geen sterke business case is voor het delen van data tussen organisaties. Op dit moment is onderzoek en innovatie het voornaamste doel voor het delen van data.

Naast de samenwerkingsmodellen uit de theorie zijn er verschillen in hoeveel partijen de data aanbieden en hoeveel partijen de data gebruiken. Het aantal aanbieders en gebruikers bepaalt de complexiteit van de samenwerking. De cases variëren van 1-to-1 (Ahold), 1-to-many (Rotterdam en Achmea), en many-to-many (energie data). Het 1-to-1 model biedt de meeste controle, terwijl de many-to-many het meest complex is. Deze complexiteit wordt mede bepaald door de openheid of juist geslotenheid van het samenwerkingsverband. Het netwerk van Ahold en Achmea zijn gesloten, terwijl het samenwerkingsverband tussen de energiepartijen en de gemeente Rotterdam juist open tot zeer open zijn.

Tabel 3. Overzicht van de eigenschappen van de big data samenwerkingsverbanden

	Ahold	**Rotterdam**	**Achmea**	**Energiepartijen**
Samenwerking	In de Ahold casus is er sprake van één partij die data deelt met meerdere partijen (**1-to-many**). Er is sprake van een **hiërarchisch**	In de Rotterdam open data casus is er sprake van één partij (gemeente Rotterdam) die data met meerdere partijen deelt (**1-to-many**). Er is	Eén partij (Achmea) deelt data met meerdere partijen (**1-to-many**). Er is sprake van een **hiërarchisch** governance model, met	In de energiecasus is er sprake van meerdere deelnemers die aan meerdere partijen data leveren (**many-to-many**). Er is sprake van een

	Ahold	**Rotterdam**	**Achmea**	**Energiepartijen**
	governance model, met Ahold als dominante partij.	sprake van een **bazaar** governance model.	Achmea als dominante partij.	**netwerk** governance model, waarbij alle deelnemers zitting gaan nemen in een netwerkorganisatie.
Doel	Commercieel doel (marketing).	Innovatie / publieke waarde als doel.	Wetenschap en kennis als doel.	Innovatie / publieke waarde als doel.
Openheid van samenwerking	Het netwerk is **gesloten**.	Het netwerk is **open**.	Het netwerk is **redelijk gesloten**.	Het netwerk is **redelijk open**.

4.2 Eigenschappen van big data

Tabel 4 geeft een overzicht van de type datasets die in de cases worden gedeeld. Allereerst is er een groot verschil in de gevoeligheid t.a.v. persoonsgegevens: van de persoonlijke zorgdata van Achmea of klantgegevens van AH klanten tot de exacte locaties van bomen in Rotterdam of geaggregeerde data over energieverbruik. In alle cases valt het op dat er vooral zeer gestructureerde data (bijvoorbeeld transactiegegevens) wordt gedeeld, en maar zeer beperkt ongestructureerde data (bijvoorbeeld data verkregen vanuit social media), zoals in de Rotterdam open data casus.

Het delen en koppelen van data gebeurt nog niet op grote schaal. Dat betekent ook dat er nog relatief weinig samenwerkingen zijn. Rotterdam open data deelt natuurlijk wel datasets, maar doet dit op een open manier, waarbij er geen duidelijke samenwerkingsvorm is gedefinieerd. Dit sluit het beste aan op het bazaar model. In beginsel kan iedere partij deelnemen en er is geen onderlinge afhankelijkheid.

Daarnaast is er nog weinig datamaximalisatie, waarbij meerdere datasets gekoppeld worden, dus de onvoorspelbaarheid van toepassingen is nog relatief beperkt. Alleen in het voorbeeld van Achmea, en mogelijk later in de Ahold case, wordt aangegeven dat er meerdere data sets worden gecombineerd. De verwerkingen vinden daar echter sterk gecontroleerd plaats en met duidelijk van te voren bepaalde doelen (wetenschappelijk onderzoek of gericht adverteren). Dit kan in de toekomst veranderen: in alle cases wordt verwacht dat de data explosief zal toenemen.

Tabel 4 Overzicht van de eigenschappen van de data in de cases

	Ahold	**Rotterdam**	**Achmea**	**Energiepartijen**
Aggrega-tieniveau	Datasets zijn op **individueel detailniveau**: alle transactiegege-vens van AH klanten met een bonuskaart.	Datasets zijn op **geaggregeerd niveau** en bevatten geen persoonsge-gevens.	Datasets zijn **op individueel detailniveau**: alle gegevens van zorggebruik van verzekerden op individueel niveau. Data zijn wel geanonimi-seerd opslagen.	Datasets **variëren in detailniveau**: van geaggregeerde gegevens tot energiegegevens op huishouden niveau. Mogelijk in de toekomst op individueel niveau door data van slimme thermometer.
Linked data	**Geen linked data.** De transactiedata wordt mondjesmaat gekoppeld met andere CRM gegevens.	**Op termijn linked data.** Datasets zullen gekoppeld worden voor apps en op termijn ook beleidsanalyses.	**Linked data.** Via de TTP Mondriaan en ZorgTTP worden datasets gekoppeld.	**Linked data.** Het platform gaat data koppelen, zodat er nieuwe datasets ontstaan.
Data maximali-satie	**Geen datamaximalisat ie.** Gerichte analyses om persoonlijke aanbiedingen te doen.	Vanuit open data gedachtegoed is het doel **datamaximalisat ie** en dienen er juist onverwachte, innovatieve toepassingen te komen op basis van de data.	**Geen datamaximalisat ie** Redelijk duidelijk wat er met de data wordt gedaan en met welke datasets deze wordt gecombineerd: alle onderzoeks-werkzaamheden worden getoetst.	**Op termijn sprake van datamaximalisat ie**. In eerste instantie worden 6 use cases uitgewerkt en is vrij duidelijk wat . Op termijn wordt de data vrijgegeven aan app ontwikkelaars.

4.3 Privacyrisico's

Tabel 5 geeft een overzicht van de (gepercipieerde) privacyrisico's in de cases. Alle cases zijn zich sterk bewust van de privacyrisico's die verbonden zijn aan het delen van big data. In grote mate is dit bewustzijn gedreven door de aanwezigheid van het juridische kader, in het bijzonder de Wbp. In beginsel zijn veel partijen hoofdzakelijk bezig met pure compliance: het voldoen aan de wettelijke vereisten. In bepaalde gevallen, zoals bij Rotterdam open data, wordt vanwege privacy bewust niet gewerkt met persoonsgegevens. De data in de sets die open worden aangeboden bevatten geen persoonsgegevens, ook niet op geaggregeerd niveau.

Desondanks is er nog steeds het bewustzijn dat ook hier risico's op kunnen treden, bijvoorbeeld omdat de open data uit de BAG gekoppeld worden aan een identificerende dataset waar een partij al over beschikt. Daarmee worden de data alsnog persoonsgegevens en is de Wbp van toepassing. De verantwoordelijkheid ligt in dat geval bij de partij die de koppeling maakt, omdat het voor die partij persoonsgegevens zijn, niet voor de gemeente Rotterdam. Het risico bij vrijgeven van een dataset over de fietspaden in de stad is bijvoorbeeld veel lager of feitelijk afwezig.

Naast het risico van (her)identificatie is nog een aantal vereisten uit het wettelijk kader prominent aanwezig. Het verkrijgen van voorafgaande (geïnformeerde) toestemming is vaak lastig. In het geval van zorgverzekeraars wordt zelfs aangegeven dat het verkrijgen van die toestemming dusdanig complex is en zoveel problemen met zich meebrengt dat de kosten daarvan hoger zullen zijn dan de te verwachten opbrengsten. Ook hier wordt dus terughoudend te werk gegaan. Het gebruik van gegevens voor wetenschappelijke doeleinden, zoals statistische analyses, vindt wel plaats. Verdergaande initiatieven, waarbij niet zozeer gehandeld wordt vanuit het oogpunt van een maatschappelijk belang, blijven vooralsnog achterwege. Dit heeft ook te maken met het feit dat het om medische gegevens gaat, die in de Wbp zijn aangemerkt als bijzondere persoonsgegevens. Daarbij geldt een zwaarder beschermingsniveau en is het dus extra gevoelig om de gegevens te verwerken voor doeleinden die vooraf niet geheel duidelijk zijn. Op dit moment wordt de grondslag vaak nog gebaseerd op het accepteren van algemene voorwaarden, waarmee toestemming gegeven wordt. Men is zich echter wel bewust dat dit geen elegante manier is. Afhankelijk van de relatie tussen de zorgpartij en het individu en het verband waarin de algemene voorwaarden geaccepteerd worden zal de juridische houdbaarheid hiervan ook beperkt zijn.

Opvallend genoeg wordt doelbinding door de respondenten niet als een probleem ervaren. En dat terwijl dit in de theorie als een van de belangrijkste problemen wordt geïdentificeerd. De toepassing van big data is niet altijd vooraf duidelijk, maar wordt gaandeweg het datamining proces ontdekt. Van te voren een concreet doel vaststellen en communiceren is daarom niet mogelijk. De verklaring voor het achterwege blijven van doelbinding als probleem in de use cases ligt wellicht in de grote terughoudendheid waarmee momenteel nog te werk gegaan wordt. Zoals aangegeven in de vorige paragraaf, zijn toepassingen vooraf duidelijk in de cases en er wordt dus gericht gewerkt naar een bepaald doel.

Een punt dat wel vaker terugkomt is de onduidelijkheid omtrent rechten en plichten bij het delen van data. Dit wordt vooral vanuit de samenwerkingsvorm geregeld en wordt daarom later behandeld.

Tabel 5 Overzicht van de privacyrisico's in de cases

	Ahold	**Rotterdam**	**Achmea**	**Energiepartijen**
Privacyrisico	**Hoog.** De data bevat (ongevoelige) persoonsgege-vens.	**Laag.** De data bevat geen persoonsgege-vens. Bij verdere koppeling van data is er een klein risico tot heridentificatie.	**Hoog.** Data bevat gevoelige persoonsgege-vens.	**Middel.** Slimme meter data bevat informatie op het niveau van zes huishoudens.
Openheid van data	De data is **gesloten** en alleen toegankelijk voor een externe analyse partij.	De data is **open** en gratis toegankelijk voor partijen buiten het netwerk.	De data is **semi-gesloten** en slecht toegankelijk voor deelnemers buiten het netwerk.	De data is **semi-open**: het is nu nog moeilijk om toegang te krijgen via data contracten, maar vanuit open data gedachtegoed zal het makkelijker worden.
Doelbinding	De data worden gebruikt voor het doel waar ze verzameld zijn, namelijk marketing.	Er is geen vaststaand doel voor de data.	Een commissie toetst of de data door gebruikers worden gebruikt conform het doel (namelijk wetenschap)	Er is geen vaststaand doel voor de data.

4.4 Eigenaarschap en accountability

Tabel 6 geeft een overzicht van eigenaarschap in de cases. In de huidige praktijk is het eigenaarschap van data vaak nog onduidelijk. In sommige gevallen wordt expliciet de vraag gesteld of individuen wellicht eigenaar zijn van hun eigen gegevens. In andere gevallen wordt er simpelweg vanuit gegaan dat de data van het bedrijf zijn dat de data heeft. Hier wordt geredeneerd vanuit de beschikking over de data en de mogelijkheid om er iets mee te doen. De meeste vragen rijzen in het zorgdomein, mogelijk omdat het daar bijzondere persoonsgegevens betreft. In de Ahold case stelt Ahold zelf eigenaar van de data te zijn. Alle verwerkingen en koppelingen vinden ook plaats in opdracht van Ahold. Er is dus duidelijk sprake van een hiërarchische relatie waarbinnen de verwerkingen plaatsvinden. Ahold is daarmee in ieder geval de verantwoordelijke voor de verwerking. Omdat Ahold opdracht geeft voor de verwerkingen stelt het ook eigenaar te zijn. Dit is ook de meest gangbare praktijk. Er is dus meer dan gemiddeld sprake van voorzichtigheid en er wordt daarom ook meer nagedacht over wie welke rechten precies heeft ten aanzien van de data. Daarnaast speelt de discussie over eigenaarschap op twee niveaus: 1) tussen organisaties die data delen en 2) tussen organisaties en individuen, ook wel user empowerment. User empowerment is een begrip waar alle

cases over nadenken. Theoretisch gezien zou user empowerment goed zijn voor vertrouwen e.d. in data diensten, maar praktisch verslechtert de business case voor big data initiatieven, ook niet-commerciële initiatieven.

In alle gevallen speelt de vraag wie eigenaar is van gekoppelde data. Als er een nieuwe data set ontstaat, wie is dan eigenaar? Of zijn de partijen die de gegevens aanleveren gezamenlijk eigenaar? Tot nog toe wordt hier pragmatisch mee omgegaan. Het probleem is wel zichtbaar, maar er wordt onderling afgestemd wie wat mag doen met data. Het komt ook voor dat er simpelweg niks wordt afgesproken. Zolang er geen conflicten optreden gaat dit natuurlijk goed.

Tabel 6 Overzicht van eigenaarschap van de data in de cases

	Ahold	**Rotterdam**	**Achmea**	**Energiepartijen**
Wie ziet zich als eigenaar?	**Eigenaarschap blijft bij de dominante organisatie.** Ahold ziet zich als eigenaar van de data.	In het open data model is er **geen eigenaarschap**. In de praktijk is de community informeel eigenaar van de data.	**Eigenaarschap blijft bij de dominante organisatie.** Na vijf jaar moeten de onderzoekers de data verwijderen.	Eigenaarschap tussen de netwerkorganisaties blijft bij de **individuele organisaties.** Eigenaarschap wordt vastgelegd in data contracten.
Toestemming	De bonuskaarthouder geeft **expliciet toestemming** tot gebruik van zijn data.	Er is **geen toestemming** nodig van burgers.	Patiënt geeft **impliciet** via de voorwaarden van de zorgverzekering **toestemming** tot gebruik van zijn data.	**Onduidelijk** in hoeverre burgers toestemming moeten geven over data op verschillende detailniveaus.
Niveau van discussie	De voornaamste discussie over eigenaarschap speelt **op organisatie-individu niveau.**	Door het open en geaggregeerd karakter van de data is er **geen discussie** over eigenaarschap.	De voornaamste discussie over eigenaarschap speelt **op organisatie-individu niveau.**	De voornaamste discussie over eigenaarschap **speelt tussen de organisaties.** Het project ziet een rol voor de gebruikers van de data om transparantie te bieden aan burgers.

Tabel 7 laat zien hoe de cases verschillen in hoe accountability oftewel aansprakelijkheid is ingericht. Sommige partijen, zoals Ahold, richten accountability sterk vanuit een compliance perspectief in. Dat betekent dat zij interne en externe

audits houden, en toezien op de naleving van de Wbp. Een deel wordt ook afgedekt door de hiërarchische relatie, waardoor relatief veel controle gehouden wordt over de data en wat ermee gebeurt. Ahold blijft verantwoordelijk en het bedrijf Symphony EYC werkt in opdracht van Ahold. Er wordt bewust niet met andere partijen samengewerkt of gedeeld. Stap voor stap worden er wel meer koppelingen gemaakt tussen datasets, maar dit is een moeizaam proces. Legacy problemen, met name bij systemen die continu moeten blijven draaien, zorgen ervoor dat dit geen eenvoudig proces is. Accountability wordt bij de zorgcasus niet expliciet genoemd. Er zijn echter wel enkele dingen daarop gericht, zoals het gebruik van Trusted Third Parties (TTPs), aandacht voor de legitieme grondslag voor gegevensverwerkingen, en verantwoorden van verwerkingen op basis van maatschappelijke doeleinden, zoals wetenschappelijk onderzoek. Bij de energiesector is er meer expliciet aandacht voor accountability. Er is duidelijk een wens voor meer transparantie, informatievoorziening en user empowerment. Hier zie je dus ook dat accountability breder wordt ingestoken en ook een belangrijke component heeft waarin de gebruiker meer inzicht en controle krijgt.

Naast accountability afspraken (bijv. de data agreements in het TEI project) of organen (bijv. TTP en toetsingscommissie bij de Achmea Health Database) speelt vertrouwen tussen partijen ook een rol. Zo geeft Achmea aan dat uiteindelijk onderzoekers worden vertrouwd in hun wetenschappelijk integriteit.

Tabel 7 Overzicht van accountability van de data in de cases

	Ahold	**Rotterdam**	**Achmea**	**Energiepartijen**
Niveau van discussie	Accountability speelt alleen tussen **organisatie en individu**, omdat er geen data wordt gedeeld tussen organisaties.	Accountability speelt wegens open data model **beperkt** een rol.	Accountability speelt voornamelijk op **interorganisatio-neel** niveau: tussen Achmea en de organisatie die de data gebruiken.	Accountability speelt zowel op **inter-organisationeel** niveau als op **organisatie-individu** niveau.
Inrichting tussen organisaties	Nvt.	Nvt.	Op inter-organisationeel niveau wordt er gebruik gemaakt van een Trusted Third Party en strikte afspraken tussen aanbieder en gebruiker van de data.	Op inter-organisationeel niveau worden nu afspraken vastgelegd op met case-based contracten. Het data platform zal op termijn door een externe partij worden geaudit.
Inrichting tussen organisatie en gebruiker	Ahold vraagt toestemming aan klanten, er wordt een anoniem alternatief aange-boden, en	Nvt.	Achmea vraagt impliciet toe-stemming via voorwaarden van verzekering. Verder is vanuit	Vooralsnog ziet het project een rol gelegd bij de gebruikers van de data om transparantie te

	Ahold	**Rotterdam**	**Achmea**	**Energiepartijen**
	klanten kunnen hun persoons-gegevens laten verwijderen uit de database van Ahold.		kostenoverwegingen de accountability naar patenten toe beperkt.	geven aan burgers / energie-consumenten.

5 Conclusie

In dit onderzoek staat de vraag centraal hoe bij het samenwerken rond big data eigenaarschap en accountability zo ingericht kunnen worden dat privacyrisico's geminimaliseerd worden. Duidelijke richtlijnen t.a.v. eigenaarschap en accountability, bovenop de Wbp, kunnen het vertrouwen van samenwerkende partijen en burgers over het verzamelen, verwerken en toepassen van data vergroten. De onderzoeksvraag is opgedeeld in vier onderdelen: 1) samenwerkingsvorm, 2) eigenschappen van de data, 3) privacyrisico's en 4) eigenaarschap en accountability. Om de onderzoeksvraag en haar onderdelen te beantwoorden is een onderzoekraamwerk ontwikkeld op basis van governance en dataprocestheorieën, en zijn vier praktijkcases (retail, gemeente, zorg en energie) onderzocht door middel van desk research en interviews. Bij de keuze van de cases stonden verschillende governance vormen centraal. In een cross-case analyse zijn de cases met elkaar vergeleken en conclusies getrokken.

Samenwerkingsvormen

- Het was niet mogelijk om een sterk marktmodel te vinden voor het delen van data. De business case is nu nog onduidelijk, mede door het gebrek aan beleidsrichtlijnen over eigenaarschap en accountability.
- Het gebrek aan markt werd geïllustreerd door het doel in de cases: organisaties richten zich op exploratie (bijvoorbeeld innovatie) in plaats van exploitatie.
- Het aantal data-aanbieders, gebruikers en de type data bepalen in grote mate de openheid en complexiteit van datasamenwerkingen. De complexiteit van samenwerking bepaalt mede hoeveel coördinatie, en dus ook eigenaarschap en accountability, nodig zijn in een samenwerking.

Eigenschappen van data

- Het is de vraag in hoeverre er gesproken kan worden over big data. Op dit moment is de data overzichtelijk, gestructureerd en zijn de toepassingen op basis van de data voorzienbaar. Er is dus nog geen sprake van datamaximalisatie en overschrijding van doelbinding bij het verwerken van de data.
- Hoe lager het detailniveau van de data, des te opener de data tussen partijen wordt gedeeld. De waarde van de data om te delen is echter hoger wanneer het detailniveau hoog is.
- Echter, in alle domeinen neemt de hoeveelheid data sterk toe en zal er steeds vaker sprake zijn van big data.
- Hetzelfde geldt voor het koppelen van data (zogenaamde linked data). Deze koppeling gebeurd nog sporadisch, maar de verwachting is dat het koppelen toeneemt.

Privacyrisico's

- Er wordt goed nagedacht over de privacyrisico's van het delen van data, zelfs waar het open data betreft. De koppeling van datasets kan immers weer leiden tot heridentificatie van personen. In dit geval is degene die de datasets koppelt verantwoordelijk voor naleving van de Wbp. Op dit moment ligt de nadruk op het precies naleven van de Wbp ("compliance"). Een reden kan de afweging zijn van marktpartijen tussen de kosten voor het inrichten van privacy en de baten die uit de data worden gehaald.

- Door het beperkt koppelen van datasets is doelbinding, zoals gesteld in de Wbp, geen probleem in de praktijk. De beweging naar datamaximalisatie en het koppelen van datasets komt duidelijk naar voren in de cases, en kan dus later tot problemen leiden.

Eigenaarschap en accountability

- Eigenaarschap is onduidelijk in de cases. De discussie van eigenaarschap speelt op twee niveaus: 1) tussen organisaties die data met elkaar delen, en 2) tussen organisaties en de personen waar de data overgaat (user empowerment).
- Op netwerkniveau blijven organisaties eigenaar over de data die ze delen met andere partijen. De onduidelijkheid over eigenaarschap ontstaat voornamelijk wanneer datasets van verschillende partijen met elkaar worden gekoppeld tot een nieuwe dataset.
- Theoretisch gezien zou user empowerment goed zijn voor het vertrouwen in data diensten, maar praktisch verslechtert het de business case voor big data initiatieven, ook niet-commerciële initiatieven.
- Accountability over de data speelt eveneens op het niveau van organisaties en personen. In de praktijk wordt accountability ingericht door middel van interne en externe audits, afspraken, en coördinatie door een derde partij (Trusted Third Party) of een nieuwe opgerichte netwerkorganisatie.
- Binnen een netwerk speelt vertrouwen tussen de partijen een belangrijke rol in het oplossen van conflicten. Dit heeft een positief effect op het beheersen van reputatierisico's, maar kan negatieve gevolgen hebben door het niet in de publiciteit brengen van privacyschendingen.

Dit onderzoek laat zien dat organisaties die data met elkaar gaan delen goed moeten nadenken over 1) hoe ze samenwerken 2) welke data ze met elkaar delen 3) welke privacyrisico's daar aan verbonden zijn en 4) hoe ze naast het naleven van de Wbp, eigenaarschap en accountability kunnen inrichten om risico's te beheersen. In het licht van een sterke toename van data en een strikter juridisch kader, neemt het belang van een afgewogen strategie waarin de gebruiker centraal staat toe. Op deze manier kan het vertrouwen in een zorgvuldige omgang met data zowel tussen organisaties als tussen organisaties en burgers toenemen.

6 Discussie

Dit rapport laat zien dat er economische kansen bestaan voor het breed toepassen van big data. Het type toepassingen hangt samen met de samenwerkingsvorm waarin data gedeeld worden tussen partijen. Momenteel zijn er echter nog twee belangrijke barrières: het business model is niet altijd duidelijk en er ligt een streng juridisch kader dat niet aansluit op de praktijk van big data toepassingen.
Met betrekking tot het business model kunnen wij hier niet alle oplossingen bieden. Wel kunnen we constateren dat in beginsel iedereen het erover eens is dat veel data waarde hebben en in waarde kunnen groeien wanneer er nieuwe toepassingen worden gevonden. De latente aanwezigheid van waarde zou tot innovatie moeten leiden. Dit vindt echter nog slechts zeer beperkt plaats en de reden daarvoor lijkt, tenminste deels, te liggen in de juridische kaders en onzekerheid die daaruit voortvloeit. Kort gezegd hindert de barrière van het juridische kader big data innovatie, waardoor de waarde uit big data toepassingen nog niet of slechts minimaal verzilverd wordt.

Hoewel het juridische kader een barrière lijkt te vormen, erkennen wij zeker het belang van goede privacybescherming van de consument. Profiling kan nadelige gevolgen hebben en de risico's lijken in het geval van big data alleen maar sterker aanwezig te zijn. De Wbp biedt echter vooral kaders voor het proces van verwerking van persoonsgegevens en mist daarmee het bredere privacyperspectief en, belangrijker nog, de mogelijke impact van gegevensverwerkingen waar consumenten daadwerkelijk door geraakt worden. Daarom is vooral user empowerment erg van belang. Dit kan o.a. via:

- het bieden van waarborgen in het geval van privacyschending;
- handhaving; en
- bewustwording van consumenten/burgers.

Voor al deze oplossingsrichtingen is het inrichten van accountability van belang (als middel om die andere zaken te realiseren). Accountability draagt dan bij aan het vertrouwen in big data toepassingen, doordat specifiek aandacht wordt besteed aan de gevolgen die op kunnen treden voor individuele consumenten. We bedoelen met accountability dat organisaties verantwoording af kunnen leggen over de gegevens die ze verwerken, op welke wijze ze dat doen, en welke gevolgen ze aan gegevensverwerkingen verbinden. De keuzes en handelingen van organisaties moeten traceerbaar worden om verantwoording over het handelen af te kunnen leggen. Deze accountability gaat vooral richting consumenten en klanten omdat het over hun data gaat die geprofiled wordt en omdat de beslissingen die daaruit volgen hen raken. Accountability is daarmee een noodzakelijke voorwaarde om de hierboven genoemde oplossingen voor user empowerment mogelijk te maken.

In het kader van big data is het echter wel belangrijk om te kijken hoe accountability praktisch vormgegeven kan worden. Vanwege de aard van big data en de onverwachte resultaten die big data analyses kunnen opleveren, is het van belang om verschillende onderdelen waarover een organisatie verantwoording kan afleggen te onderkennen:

- **Proces**: de volgorde waarin persoonsgegevens worden verwerkt. Inzicht hierin is nodig om te bepalen welke partijen welke gegevens verwerken. Dit onderdeel

wordt afgedekt door de Wbp. Bovendien wordt hiermee ook duidelijk of gegevens(sets) gedeeld zijn met of afkomstig zijn van andere partijen. Dit is van belang omdat de juridische begrippen van verantwoordelijke en bewerker in de huidige praktijk vaak diffuus zijn.

- **Eigenaarschap**: welke partij heeft zeggenschap over de data. Duidelijkheid hierover verschaft inzicht in welke partijen geautoriseerd zijn om persoonsgegevens te verwerken. Er wordt duidelijkheid verkregen over de zeggenschap over data en of de set wel gebruikt mocht worden door de desbetreffende partij en onder welke voorwaarden. Op basis van het proces kan bepaald worden welke partijen relevant zijn in het kader van eigenaarschap.

- **Dataset**: individuele (typen) gegevens per dataset. Inzicht hierin geeft duidelijkheid in de gegevens die zijn gebruikt en waar deze gegevens vandaan komen. Belangrijke aspecten hierbij zijn of het oorspronkelijk ook al om persoonsgegevens ging of niet, of de gegevens door de consument zelf zijn verstrekt, of het gaat om geobserveerde (gedrags)gegevens, of om afgeleide data, en of zorgvuldigheid is betracht.

- **Algoritme**: de verwerking van de persoonsgegevens. Algoritmes maken duidelijk op welke wijze data zijn verwerkt en gecombineerd. Welke algoritmes worden gebruikt en zijn de bewerkingen reproduceerbaar? Hierbij is het ook van belang om na te gaan of er een bepaalde bias in de algoritmen of in de uitkomsten zitten. De oorsprong van de dataset en de gegevens in de dataset kunnen daar ook inzicht in geven.

De consument of klant zou in het verlengde van de hierboven genoemde richtingen ook handvatten moeten krijgen om accountability af te dwingen. Een proactieve houding van organisaties en samenwerkingsvormen (al dan niet daartoe aangezet vanuit de overheid) verdient echter wel de voorkeur. Het is immers vaak een relatief grote stap voor consumenten om hun inzagerechten etc. uit te oefenen. Bovendien zou actie vanuit de organisaties leiden tot grotere bewustwording van de processen en beslissingen waar zij mee bezig zijn, wat al bij zou dragen aan het verbeteren van de positie van de consument en klanten. Bovenstaande analyse geeft een beeld van de mate van controle over data die verschillende organisatievormen kenmerken en helpt organisaties om de reikwijdte van privacyaspecten in big data toepassingen te overzien.

Het inrichten van accountability bij big dat toepassingen en heldere afspraken over eigenaarschap van data kunnen helpen om het vertrouwen in big data toepassingen te vergroten. Omdat de focus ligt op organisatorische en technische maatregelen, waarmee een organisatie verantwoording aflegt over de volledige wijze waarop met data omgegaan wordt en welke waarborgen en remedies zijn ingebed om de belangen van consumenten te beschermen, krijgen organisaties een concreter beeld van de belangrijkste aandachtspunten bij het ontwikkelen en uitrollen van nieuwe diensten, gebaseerd op big data. Deze helderheid geeft richting aan innovatie en ondersteunt nieuwe initiatieven. Daarmee biedt het handvatten om nieuwe diensten te ontwikkelen, zonder dat vooraf een halt wordt toegeroepen op grond van formele juridische vereisten. Bovendien betekent de aandacht voor consumentenbelangen en het bieden van waarborgen en remedies dat de belangen van de consument uiteindelijk mogelijk beter gewaarborgd worden

dan op grond van de vereisten uit de Wbp. De praktische implementatie binnen organisaties moet uiteraard zo georganiseerd kunnen worden, dat bedrijven niet slechts aan extra regels moeten voldoen, maar dat deze ook daadwerkelijk het delen van gegevens ondersteunt.

Met een accountability benadering krijgen organisaties de ruimte om innovatieve diensten te ontwikkelen. Tevens zijn er meer mogelijkheden om te experimenteren met big data. Een helder kader voor eigenaarschap helpt bovendien om rechten en aansprakelijkheden duidelijk te beleggen. Daarmee wordt ook duidelijk wie wat mag doen met bepaalde data sets. In het bijzonder in het geval van linked data, waar een grote winst te behalen valt voor big data toepassingen, is dit erg behulpzaam. De huidige twijfel en onzekerheid hinderen verdere ontwikkelingen. Experimenteren wordt nagenoeg niet gedaan, waardoor de toegevoegde waarde van big data nog niet verzilverd kan worden. Een benadering waarbij de aandacht ligt op bescherming van de consument waar deze geraakt wordt door gevolgen van big data toepassingen is daarom mogelijk vruchtbaarder dan de huidige benadering die de Wbp voorschrijft.

7 Referenties

- boyd, d. en K. Crawford (2012). Critical Questions for Big Data, *INFO. COMM. & SOC'Y* (MAY 2012), p. 6.
- Demil, B., en X. Lecocq (2006). Neither market nor hierarchy nor network: The emergence of bazaar governance. *Organization studies* 27.10: 1447-1466.
- El Emam, Khaled, et al. (2012) De-identification methods for open health data: the case of the Heritage Health Prize claims dataset. *Journal of medical Internet research* 14.1.
- Huijboom, N.M., en T.A. Van den Broek. (2011) Open data: an international comparison of strategies. *European Journal of ePractice* 12.1, 1-13.
- Koot, M. R. (2012). Measuring and predicting anonymity. Proefschrift, UvA.
- Kumar, K., en H.G. Van Dissel (1996). Sustainable collaboration: managing conflict and cooperation in interorganizational systems. *Mis Quarterly*: 279-300.
- Lowndes, V., en C. Skelcher (1998). The dynamics of multi-organizational partnerships: an analysis of changing modes of governance. *Public administration* 76.2: 313-333.
- Ministerie van Economische Zaken (2013). Kabinetsvisie op e-privacy: op weg naar gerechtvaardigd vertrouwen, http://www.rijksoverheid.nl/documenten-en-publicaties/kamerstukken/2013/05/24/kamerbrief-met-kabinetsvisie-op-e-privacy.html.
- Narayanan, A. en V. Shmatikov (2008). Robust De-anonymization of Large Sparse Datasets, *2008 IEEE Symposium on Security and Privacy*, p. 119.
- O'Toole Jr, L.J. (1997). Treating networks seriously: Practical and research-based agendas in public administration. *Public administration review*, 45-52.
- Powell, W. Neither market nor hierarchy. *The sociology of organizations: classic, contemporary, and critical readings* 315 (2003): 104-117.
- Provan, K.G., en P. Kenis. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of public administration research and theory* 18.2 (2008): 229-252.
- Schreuders, E. (2001). Data mining, de toetsing van beslisregels & privacy, *ITER* 48, 2001, p. 30.
- Sweeney, L. (2002). K-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), 557-570.
- Tene, O. en J. Polonetsky. Big Data for All: Privacy and User Control in the Age of Analytics, p.22. Beschikbaar via SSRN: http://ssrn.com/abstract=2149364.
- White, T. (2012). *Hadoop: The Definitive Guide.* O'Reilly Media, p. 3. ISBN 978-1-4493-3877-0.

8 Annex A: Interview protocol

1. **Persoonlijke introductie**

2. **Introductie BTK project**
   - Aanleiding van het project
   - Doel van het project
   - Korte uitleg over wat wordt verstaan onder big data samenwerkingen

3. **Introductie interview**
   - Onderwerpen van het interview: de samenwerking, de eigenschappen van de big data, privacyrisico's en eigenaarschap & accountability
   - Lengte van het interview
   - Omgaan met data van het interview

4. **Vragen**

a. De big data samenwerking
   - Kunt u wat meer vertellen over het big data samenwerkingsverbanden waarin u betrokken bent?
   - Hoe is het samenwerkingsverband tot stand gekomen (per samenwerkingsverband)?
   - Wat is het doel van de samenwerking?

   Bijvoorbeeld: innovatie, efficiëntie, delen van risico's of middelen, nieuwe diensten, intelligence, wetenschap.

   - Hoe heeft dit doel invloed op de manier van delen van data?
   - Welke en hoeveel partijen werken samen?
   - Wat is de relatie tussen de samenwerkende partijen? Hoe vaak hebben de partijen eerder samengewerkt?
   - Door wie en hoe wordt de samenwerking gecoördineerd?
   - In hoeverre is de samenwerking open voor nieuwe partijen? Onder welke voorwaarden is dit mogelijk?
   - Hoe zou u de samenwerking willen inrichten? Welke belemmeringen worden daarbij ervaren?

b. De eigenschappen van de big data
   - Kunt u de data van uw samenwerking beschrijven?
   - In hoeverre bevat de data persoonsgegevens?
   - In hoeverre is er een vaststaand doel voor het verzamelen, verwerken en toepassen van data? In hoeverre kan dit doel wijzigen?
   - Welke toepassingen van de data voorziet u en de andere partijen?
   - Hoe en door wie wordt de data verzameld, verwerkt en toegepast? (mag 1 voor 1 worden gevraagd)
   - Hoe wordt de data technologisch gezien gedeeld?

   Bijvoorbeeld: centrale database, peer-to-peer transacties, etc.

- Welke barrières ervaar je bij het delen van data met andere partijen? Welke oplossingen zijn hiervoor?

c. Privacyrisico's
   - Wie is verantwoordelijk voor de naleving van de Wet Bescherming Persoonsgegevens?
   - Hoe wordt aan de vereisten van de Wet Bescherming Persoonsgegevens voldaan bij het verzamelen, verwerken en het toepassen van de data? (mag 1 voor 1 worden gevraagd)
   - In hoeverre zijn de partijen zeker dat de datasets geanonimiseerd zijn en blijven?
   - Terugkomend op de barrières, in hoeverre vormen privacyrisico's een belemmering voor samenwerken op het gebied van big data?
   - Zijn er aspecten van de WBP die volgens u anders ingericht kunnen worden? Wat zijn hiervan de voordelen of de risico's?

d. Eigenaarschap en accountability
   - Wie is eigenaar van de data?
   - Wie ziet toe op de rechtmatige verzameling, verwerking en toepassing van de data? (mag 1 voor 1 worden gevraagd)
   - Hoe wordt er juridisch en technisch toegezien op de rechtmatige verzameling, verwerking en toepassing van de data? (mag 1 voor 1 worden gevraagd)

   Bijvoorbeeld juridisch: vertrouwen, contracten, licensing, sancties / boetes, etc.
   Bijvoorbeeld technisch: vastleggen van verkeer, access control, metadata, watermarking, etc.

   - Hoe houdt u zelf controle over het gebruik van de data?
   - Wie is accountable bij eventuele conflicten of schandalen?
   - Wat zijn de afbreukrisico's van de big data samenwerking voor u?
   - Hoe ziet u het eigenaarschap idealiter georganiseerd? Wat moet hiervoor veranderen?

5. Afsluiting
   - Bedankt!
   - Vervolg van het project en toezending van het eindrapport.
   - Evt. wie nog meer te interviewen in het netwerk (m.n. toevoegen / rijkere data)?

9 Annex B: lijst met geïnterviewde personen

- Barry Egberts (Achmea)
- Ferry de Groot (gemeente Rotterdam)
- Jan-Peter Larsen (Sense-OS)
- Judith Lemmens (Hogeschool Rotterdam)
- Karin de Goederen (gemeente Rotterdam)
- Reind van Olst (2CoolMonkeys)
- Roland Tabor (Ahold)
- Thomas de Groen (Eneco)
- Willem van den Bosch (TNO)