Beleid classificatie van onderszoeksdata

Versie: 1.0, december 2024

VU VRIJE UNIVERSITEIT AMSTERDAM

vu.nl/rds

Inhoud

Waarom deze notitie?

Onderzoekers genereren, verkrijgen, openen, delen en verwerken onderzoeksdata. Deze data worden opgeslagen en idealiter gearchiveerd. Van onderzoekers wordt verwacht dat zij proactief zijn in de bescherming van hun onderzoekdata. Om welke beveiligingsmaatregelen gaat het dan? En wanneer moeten deze maatregelen genomen worden? Voor welke typen data en gebruik gelden deze maatregelen? En wat zijn de consequenties daarvan?

Dit beleidsdocument is bedoeld als raamwerk voor onderzoekers en ondersteuners voor het effectief en veilig omgaan met onderzoeksdata en het beantwoorden van bovenstaande vragen. Dit beleidsdocument gaat uit van bestaande juridische eisen om data die als vertrouwelijk en/of gevoelig worden beschouwd, veilig te behandelen en adequaat te beschermen tegen diefstal, verlies of ongeoorloofde toegang of gebruik. Hiermee willen we de data van onze studenten, medewerkers en (proef)personen en onze organisatie beschermen. Daarnaast willen we voldoen aan de relevante wetgeving, zoals de Algemene Verordening Gegevensbescherming[1] (de AVG).

Adequate bescherming verschilt echter per type en inhoud van de data, ofwel per classificatie daarvan. Niet alle data is gevoelig of vertrouwelijk en vereist daarmee het hoogste niveau van bescherming. Specifieke maatregelen zijn afhankelijk van de mate van bescherming die nodig is om de risico's op onrechtmatige toegang en gebruik te beperken. Efficiënt gebruik van de beschikbare financiële middelen speelt daarnaast een rol in de keuze van de beschermingsmaatregelen: de kosten moeten proportioneel zijn aan het belang van de bescherming van de data.

[1] Bepaalde typen data vallen onder specifieke wet- en regelgeving, zoals de Wet Medisch wetenschappelijk Onderzoek met mensen (WMO), *Code of conduct for medical research*, de Wet op de dierproeven, etc.

1. Classificatie

Doel van het classificeren is het maken van een inschatting van de gevoeligheid van de data, het belang van de informatie en de daarbij horende graad van beveiliging. Door onderzoeksdata te classificeren is het mogelijk om deze op passende manier te beschermen. Het gaat daarbij om de juiste mate van beveiliging die passend is bij de geïdentificeerde risico's. Dit maakt het mogelijk om te bepalen waar de data wel, of juist niet verwerkt kunnen worden, en onder welke voorwaarden.

1.1 Verantwoordelijkheid

De VU is wettelijk verplicht haar data goed te beveiligen. De eerste stap naar een goede beveiliging is het maken van een inschatting van de gevoeligheid en vertrouwelijkheid van onderzoeksdata. De verantwoordelijkheid voor het maken van de classificatie ligt bij de eindverantwoordelijke onderzoeker. De eindverantwoordelijke onderzoeker kan hierbij ondersteuning krijgen bij van de Data stewards, de Coördinatoren Informatiebeveiliging, de Privacy Officer IT, de Information Security Officers en de RDM Support Desk.

1.2 Beschikbaarheid, integriteit en vertrouwelijkheid

Vertrouwelijkheid

Hoe zorgen we ervoor dat er een goede inschatting gemaakt wordt van de mate van gevoeligheid en/of vertrouwelijkheid van de data, zodanig dat er passende beveiligingsmaatregelen genomen kunnen worden die ervoor zorgen dat deze niet onbedoeld en ongewenst toegankelijk zijn voor degenen die hiertoe niet geautoriseerd zijn?

In de basis heeft classificeren bij de VU zowel betrekking op informatie (data/informatie) als op de systemen waarin deze informatie wordt verwerkt/opgeslagen[2]. Het kader dat hiervoor gebruikt wordt is classificatie langs drie assen: *Beschikbaarheid, Integriteit en Vertrouwelijkheid*, afgekort als BIV.

Bij de classificatie van onderzoeksdata ligt vaak de grootste nadruk op de as vertrouwelijkheid. Toch kunnen ook de twee andere assen, beschikbaarheid en integriteit, van belang zijn. Het (tijdelijk) niet beschikbaar zijn van data (beschikbaarheid) of het ongeoorloofd kunnen wijzigen (integriteit) van onderzoeksdata kunnen grote consequenties hebben voor een onderzoeksproject.

[2] Bij de aanschaf, ontwikkeling, implementatie, in gebruik name van nieuwe applicaties dient er altijd een BIV-classificatie en risicoanalyse uitgevoerd te worden. De beoogde systeemeigenaar is verantwoordelijk voor het (laten) uitvoeren daarvan (bijv. via Information Security Officer, een solution architect of in gezamenlijkheid).

	Risiconiveau			
Beschikbaarheid	laag	midden	hoog	
Integriteit	laag	midden	hoog	
Vertrouwelijkheid	laag	midden	hoog	zeer hoog

Voor de assen Beschikbaarheid en Integriteit hanteren we drie niveaus: "laag", "midden", "hoog". Voor de as Vertrouwelijkheid zien we echter dat de classificatie "Hoog" niet alle onderzoekers in verschillende faculteiten bedient. Daarom hanteren we hier vier risiconiveaus: "laag", "midden", "hoog" en "zeer hoog"[3]. In een notendop, data die we als "zeer hoog" classificeren, hebben meer beschermingsmaatregelen nodig dan data die onder "hoog" vallen. Het bijbehorend onderzoek is vertrouwelijk(er), maar niet "geheim". Dit betreft vooral onderzoeksdata die persoonsgegevens bevatten, maar niet uitsluitend en niet in alle gevallen. Het onderscheid tussen "hoog" en "zeer hoog" biedt meer mogelijkheden om maatregelen te kunnen nemen, maar ook om beter passende richtlijnen met betrekking tot de maatregelen te kunnen geven (zie Bijlage 1).

Zoals eerder aangegeven worden ook applicaties geclassificeerd langs de lijnen Beschikbaarheid, Integriteit en Vertrouwelijkheid. Bijlage 3 bevat een overzicht van de belangrijkste, door de VU aanbevolen, data-management applicaties en het niveau van vertrouwelijkheid waarop deze zijn geclassificeerd.

1.3 Kaders voor het bepalen van risiconiveaus

Classificeren van onderzoeksdata en het nemen beveiligingsmaatregelen betekent in de praktijk dat we balanceren tussen beschermen en zo open mogelijk maken van data zodat kennis optimaal kan renderen. Enerzijds dient de VU te voldoen aan wet- en regelgeving en moet gevoelige data beschermd worden tegen onrechtmatig gebruik. Anderzijds streven we naar betrouwbaarheid en verifieerbaarheid van data.

De beoordeling van vertrouwelijkheid is primair verbonden aan de data zelf en niet aan het proces waarvoor de data gebruikt worden. Bijvoorbeeld, medische gegevens over onderzoeksdeelnemers zijn altijd gevoelig, ongeacht of data gedeeld worden met externe onderzoekspartners. In bepaalde gevallen moet echter ook het proces van verwerking van informatie worden meegewogen. Zo kan bijvoorbeeld het combineren van data met een lage vertrouwelijkheid leiden tot informatie met een hoge vertrouwelijkheid.

[3] Naast deze vier risiconiveaus is er ook een klasse 'Geheim' in het geval dat de gegevens staats- of militaire geheimen betreffen. Deze beschouwen we als 'buitencategorie' en is geen onderdeel van de standaard data-classificatie en daarom ook niet het doel van dit document.

1.3.1 Beschikbaarheid

Onder beschikbaarheid wordt verstaan: het waarborgen dat geautoriseerde gebruikers op de juiste momenten tijdig toegang hebben tot data en aanverwante voorzieningen. Dit geeft aan in hoeverre verlies, of tijdelijk niet bereikbaar zijn van onderzoeksdata voor de organisatie en het onderzoek acceptabel is. Elementen die beschikbaarheid bepalen zijn bijvoorbeeld een betrouwbare stroomvoorziening, betrouwbare reservekopieën, het uitsluiten van zogeheten 'single points of failure' en het bestaan van voldoende toegangsmogelijkheden voor het beoogde aantal gelijktijdig te verwachten gebruikers.

1.3.2 Integriteit

Onder integriteit wordt verstaan: het waarborgen van de correctheid en de volledigheid van data en verwerking. In hoeverre is de informatie in overeenstemming is met de werkelijkheid (correct, volledig en actueel)? Het voorkomen van risico's op de integriteit van onderzoeksdata zijn bijvoorbeeld het kunnen aanbrengen van wijzigingen alleen door geautoriseerde personen, het registreren van wijzigingen, bescherming van data tijdens transport.

1.3.3 Vertrouwelijkheid

Onder vertrouwelijkheid wordt verstaan: het waarborgen dat data alleen toegankelijk is voor degenen die hiertoe geautoriseerd zijn. In hoeverre zijn de data toegankelijk en onder welke condities. Matregelen om de vertrouwelijkheid van data te beschermen zijn bijvoorbeeld versleuteling van informatie (encryptie) en authenticatie van de gebruiker zodra deze zich toegang tot de gegevens wil verschaffen.

2. Classificatieproces

2.1 Principes

De volgende principes vormen de uitgangspunten bij de dataclassificatie:

1. Het initiatief voor het uitvoeren van een dataclassificatie ligt bij de eigenaar van de betreffende dataset, in de regel is dat de onderzoeker. Hiervoor kan een onderzoeker gebruikmaken van verschillende hulpmiddelen, zoals de data-classificatie tool[4];

2. Dataclassificatie is in principe van toepassing op alle data die door de VU verwerkt en opgeslagen worden. Het is echter niet realistisch om voor ieder onderzoek een volledige classificatie en risicoanalyse uit te voeren;

3. De toe te kennen classificatie dient reëel te zijn en in verhouding te staan tot de daadwerkelijke schade die zou kunnen ontstaan bij een mogelijke inbreuk op de data;

4. In de regel zullen de maatregelen die op applicatie-/procesniveau worden genomen volgen uit de classificatie van de data. Als een applicatie/proces een lager niveau ondersteunt dan voor (een deel van) de data nodig is, zullen er extra maatregelen getroffen moeten worden specifiek voor de hoger geclassificeerde data. Het bepalen daarvan is altijd gebaseerd op afstemming tussen de onderzoeker en een Information Security Officer en de IT Privacy Officer. Als dit niet mogelijk is, kan op basis van een risico-analyse een uitzondering worden geregistreerd die geaccordeerd moet worden door de verantwoordelijke onderzoeker, Chief Information Security Officer of in het uiterste geval het CvB[5]. Dit is in principe een (tijdelijke) risico-acceptatie waarin wordt afgesproken hoe lang het betreffende risico mag voortduren en wanneer het gemitigeerd dient te worden.

2.2 Hoofdlijnen

Classificatie van data is afhankelijk van de data zelf, het proces en het informatiesysteem waarmee de data verwerkt wordt. Het proces wordt ondersteund via vragenlijsten, onder andere in een Data Management Plan (DMP). Ten aanzien van Vertrouwelijkheid kan gebruik gemaakt worden van de research data classificatie tool en de data storage finder als hulpmiddelen.

[4] De data-classificatietool is een interactief hulpmiddel waarmee onderzoekers een basisbeoordeling van de risico's en vertrouwelijkheidsaspecten met betrekking tot het gebruik van data in een onderzoeksproject kunnen uitvoeren. De tool biedt richtlijnen voor verdere stappen die genomen moeten worden ten aanzien beschermingsmaatregelen: https://vu.nl/en/research/dataclassification.

[5] De 'chain of command' is hier: onderzoeker/PI -> afdelingshoofd -> decaan -> CvB.

Het classificatieproces bestaat in hoofdlijnen uit vier fasen:

1. **Inventarisatie**: De classificatie van de data begint met het vaststellen om welke data het gaat en welke wet- en regelgeving mogelijke eisen stelt aan verwerking en opslag. Deze stap vindt al plaats bij het opstellen van een DMP[6]. De (facultaire) Data Steward(s) en Privacy Champion(s) zijn hierbij betrokken;

2. **Bepaling van impact en niveau**: Aansluitend op de inventarisatie wordt bepaald hoe groot de kans op en de impact van inbreuken is op de aspecten Beschikbaarheid, Integriteit en Vertrouwelijkheid. Deze inschatting leidt tot de eindclassificatie voor deze aspecten. Deze eindklasse is richtinggevend voor het treffen van maatregelen. Daarbij is de as Vertrouwelijkheid voor het domein onderzoek de meest belangrijke. Algemeen kan worden gesteld dat de impact op de assen Beschikbaarheid en Integriteit in de meeste gevallen medium is;

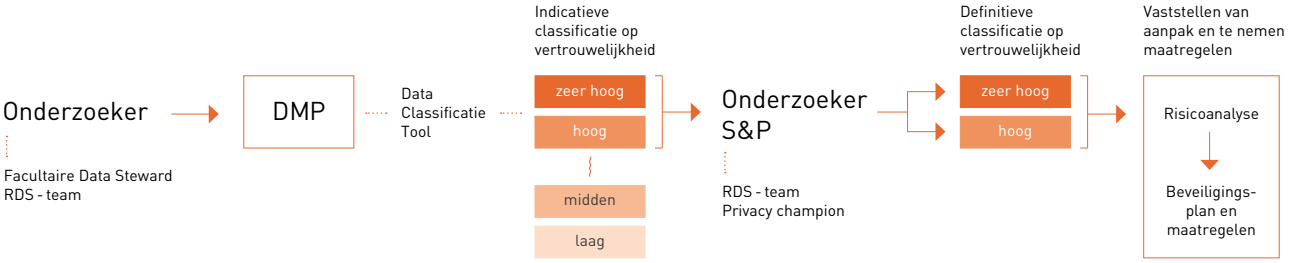
3. **Vaststellen van aanpak en maatregelen**: Afhankelijk van de classificatie en impact worden er passende maatregelen genomen. Dit betreft onder andere de keuze voor de juiste applicatie, opslaglocatie, toegangscondities, bewaartermijnen en aanvullende beveiligingsmaatregelen. Onderzoekers kunnen hierover advies krijgen van (facultaire) Data Stewards. In de regel zal dit, in geval van classificatie “hoog” en “zeer hoog”, in de vorm van een adviesgesprek worden uitgevoerd (het is raadzaam om hierbij ook een Information Security Officer en als er persoonsgegevens verwerkt worden ook de Privacy Officer IT te raadplegen). Hieruit kan volgen dat er een uitgebreide risicoanalyse nodig is waarin uitgeschreven wordt wat de mogelijke risico’s en bijpassende maatregelen zijn. Deze risicoanalyse wordt uitgevoerd door Information Security Officers met input van de onderzoeker. Naar aanleiding van de risicoanalyse stellen Information Security Officers een beveiligingsplan op. Deze vormt de basis voor het inregelen van passende maatregelen. Daar waar nodig met ondersteuning van IT voor Onderzoek of de business architecten;

4. **Beoordeling maatregelen**: In deze stap wordt gekeken of de hiervoor vastgestelde maatregelen volstaan, deze zijn geïmplementeerd, of dat er aanvullende acties nodig zijn of bepaalde restrisico’s acceptabel zijn. Dit is de taak van de Information Security Officer in samenspraak met de onderzoeker.

[6] Nog niet voor alle onderzoeksprojecten wordt er een DMP gemaakt en RDM-ondersteuning is lang niet altijd betrokken is bij de advisering en/of beoordeling van een DMP. Dat betekent dan ook dat het proces dat hier beschreven wordt uitgaat van een ideale situatie. In de praktijk is dat dus niet altijd het geval.

2.3 Aanpak

Een meer concrete invulling van de totstandkoming van een data-classificatie, een mogelijk daaruit voortkomende risicoanalyse en het bepalen en inregelen van beveiligingsmaatregelen is hieronder geschetst. Het startpunt hiervoor is een Data Management Plan aangezien de meeste onderzoekers een DMP maken en dat in veel gevallen ook verplicht wordt gesteld. Daarnaast is een DMP ook een laagdrempelige basis. Hieronder zijn de te volgen stappen geformuleerd.



Afbeelding 1. Visualisatie van het proces van classificatie en inregelen van beveiligingsmaatregelen onderzoeksdata. Zoals in de tekst is beschreven, is de onderzoeker verantwoordelijk voor het algehele proces en de Information Security Officers voor het opstellen van een risicoanalyse en beveiligingsplan. Dit in samenwerking met de onderzoeker en op basis van advisering van data management professionals.

1. Via het DMP en/of in overleg met een facultaire Data Steward wordt duidelijk dat een onderzoeker gegevens verwerkt die een Hoge of Zeer Hoge classificatie op Vertrouwelijkheid hebben (Andere reden voor het uitvoeren van een risicoanalyse (ook als de data "midden" scoren op vertrouwelijkheid) zijn bijvoorbeeld: complexe datastromen waar data van en naar verschillende systemen getransporteerd wordt, samenwerking met meerdere partners en samenwerking met partners buiten de EU). Een andere mogelijkheid is dat dit duidelijk wordt via een vraag (meestal gericht op waar data opgeslagen dient te worden) die een onderzoeker stelt aan de RDM Support Desk of een facultaire RDM-ondersteuner;

2. De RDM Support Desk of facultaire Data Steward verwijst de onderzoeker door naar de afdeling Security & Privacy, waar Information Security Officers en de Privacy Officer IT kunnen helpen een BIV uit te voeren samen met de onderzoeker. Hieruit volgt een data-classificatie. Een dergelijke BIV-classificatie wordt bij voorkeur voorafgegaan door afstemming tussen de onderzoeker en een RDM-ondersteuning. Dit om vooraf vast te stellen of een volledige BIV daadwerkelijk nodig is en om voldoende documentatie op te stellen voorafgaand aan een BIV-traject. Een volledig ingevuld DMP is daarbij nodig. Wanneer deze niet beschikbaar is dan zullen de relevante vragen uit een DMP doorlopen moeten worden;

3. Op basis van de data-classificatie wordt besloten of er een risicoanalyse nodig is. Hierin wordt in kaart gebracht welke datastromen er zijn, waar deze staan en wie erbij kan, en of data eventueel verplaatst worden. Een risicoanalyse wordt uitgevoerd met de onderzoeker, Information Security

Officers en eventueel de projectleider van en onderzoeksproject of een Data Steward of Data Manager van de betreffende faculteit of afdeling. De RDM Support Desk neemt daarbij de rol op zich van "Case Manager" en zorgt ervoor dat de juiste mensen betrokken worden bij het maken van een risicoanalyse en eventueel te nemen beveiligingsmaatregelen. Daarmee heeft een onderzoeker altijd een aanspreekpunt;

4. Op basis van de risicoanalyse wordt er een beveiligingsplan geschreven met concrete besluiten over beveiliging van de data. Daarbij wordt er onder andere gekeken naar welke beschikbare tools er gebruikt worden en welke risicomitigerende maatregelen eventueel geïmplementeerd worden. Ook hiervoor geldt dat de "Case Manager" vanuit de RDM Support Desk helpt bij het betrekken van de juiste mensen en expertise. Dit kan een functioneel beheerder van de gekozen applicatie zijn, maar ook Information Security Officers, de Privacy Officer IT, een medewerker van IT voor Onderzoek of andere benodigde expertise. Dit is sterk afhankelijk van de casus zelf, indien het "zeer hoog" gevoelige geclassificeerde data betreft is maatwerk de regel.

5. Tot slot worden conclusies en afspraken vastgelegd en worden de te nemen maatregelen doorgevoerd. Ook in deze stap is het van belang dat mensen met de juiste expertise betrokken zijn bij de uitvoering hiervan.

3. Beveiligingsmaatregelen

Het resultaat van de classificatie wordt per BIV-aspect bepaald. Op grond daarvan kan er voor ieder aspect (Beschikbaarheid, Integriteit, Vertrouwelijkheid) een classificatie vastgesteld worden en beveiligingsmaatregelen worden geadviseerd en getroffen.

De standaardmaatregelen bij de VU zijn de minimale maatregelen (baseline) die we altijd nemen. Als de classificatie op de as vertrouwelijkheid “midden” is, dan zijn de maatregelen aanvullend op de standaardmaatregelen. Bij classificatie “hoog” (en bij het aspect Vertrouwelijkheid “zeer hoog”) zullen er verdere aanvullende maatregelen nodig zijn.

Om niet iedere keer opnieuw te moeten beoordelen welke maatregelen getroffen moeten worden is er een matrix opgesteld waarin iedere classificatie automatisch gekoppeld wordt aan een set minimale maatregelen. Deze maatregelen zijn op hoofdlijnen gedefinieerd en laten ruimte voor gedetailleerde invulling. Niet alle te nemen maatregelen zijn al operationeel en ook niet alle te nemen maatregelen bestaan uit maatwerkoplossingen.

Hoewel veel maatregelen technisch van aard zijn, dient niet uit het oog verloren te worden dat de wijze waarop gebruikers omgaan met data en informatie minstens zo belangrijk, zo niet belangrijker, is dan de technische maatregelen die we kunnen treffen. Tegen onverantwoord gedrag van gebruikers is geen technische maatregel opgewassen.

Bijlage 1: Risiconiveaus

Beschikbaarheid

Klasse	Beschikbaarheid van informatie	Voorbeelden
Laag	Data kunnen langere tijd, ruim een week, niet beschikbaar zijn Verlies van data is geen ramp omdat deze eenvoudig gereproduceerd kunnen worden	Publiek beschikbare data, software en hardware
Midden	Herstel van gegevens na gegevensverlies mag niet langer dan een week duren op voorwaarde dat ze daarna wel weer volledig beschikbaar zijn en onbeschikbaarheid niet meer dan een licht negatief effect heeft op onderzoeksdoelen en reputatie van onderzoeksgroep, faculteit of VU De data kunnen maximaal een dag niet beschikbaar zijn	Data en software die binnen korte tijd vervangbaar zijn
Hoog	Data mogen niet langer dan een uur niet beschikbaar zijn	Patiëntendata Analysedata

Integriteit

Klasse	Verandering t.a.v. correctheid en volledigheid	Voorbeelden
Laag	Verandering van de informatie is toegestaan Kleine fouten in data en analyses is toegestaan	Onderzoeksgegevens die met een zekere mate van gebrekkigheid zijn aangeleverd Informatie aangeleverd via enquêtes
Midden	Minimale fouten zijn acceptabel en hebben minimale negatieve invloed op onderzoeksdoelen, veiligheid, budget of reputatie	Onderzoeksgegevens met een zeer beperkte tolerantie voor foutieve vermeldingen en impact op de resultaten
Hoog	Fouten in gegevens of berekeningen zijn niet toegestaan, aangezien dit aanzienlijke negatieve gevolgen voor onderzoeksdoelen, veiligheid, gezondheid, budget of reputatie kan hebben Verlies van gegevens mag op geen enkel moment gebeuren, omdat dit een aanzienlijke negatieve impact heeft op de doelen van het onderzoek, veiligheid, gezondheid, budget of reputatie	Biomedisch onderzoek High-risk chemische experimenten Patiëntendossiers

Vertrouwelijkheid

Klasse	Toegang tot gevoelige/vertrouwelijke gegevens	Voorbeelden
Laag	De gegevens zijn bedoeld voor openbaarmaking Het verlies van de vertrouwelijkheid van de gegevens heeft geen nadelige gevolgen op onderzoeksdoelen, veiligheid, budget of reputatie	Onder een open licentie gepubliceerde onderzoekdata/ software Zeer variabele fysieke metingen, bijvoorbeeld bloeddruk, hartslag, bloed-glucose, lichaamstemperatuur Gecodeerde kwalitatieve gegevens Samenvattende statistieken Publieke domein historische data Gecodeerde statistische data
Midden	De gegevens mogen alleen beschikbaar zijn voor een specifieke groep Het verlies aan vertrouwelijkheid van de gegevens kan lichte negatieve gevolgen hebben voor onderzoeksdoelen, veiligheid, budget of reputatie	IP- en MAC-adressen van onderzoeksonderwerpen Ruwe vragenlijstgegevens van niet-kwetsbare onderwerpen met demografische informatie Vragenlijstgegevens over gevoelige onderwerpen en/of kwetsbare personen die zijn verwerkt om heridentificatie moeilijker te maken Video-opnamen met wazig gemaakte gezichten en aange-paste stemmen Transcripten van interviews waarin de identificerende infor-matie is vervangen door pseudoniemen Defaced neuroimages van kwetsbare personen
Hoog	De gegevens mogen alleen beschikbaar zijn voor een specifieke groep De gegevens bevatten gevoelige informatie over individuele personen Bescherming is vereist door wet-/regelgeving en contextafhankelijk Het verlies aan vertrouwelijkheid van de gegevens zou een aanzienlijke negatieve gevolgen hebben voor onderzoeksdoelen, veiligheid, budget of reputatie Het gebruik van deze gegevens kan een negatieve impact hebben op iemands leven.	Gepseudonimiseerde strafdossiers van minderjarigen Bestanden met namen en contactgegevens van proefper-sonen Gegevens met geboortedatum en 6-cijferige postcode van proefpersonen Onderzoeken waar de functie binnen een organisatie van deelnemers belangrijk is makkelijk te achterhalen is Video observaties van spelende kinderen Ruwe vragenlijstgegevens over gevoelige onderwerpen Ruwe vragenlijstgegevens van kwetsbare proefpersonen met gedetailleerde demografische informatie Genetische gegevens van niet-kwetsbare proefpersonen
Zeer hoog	De gegevens mogen niet of onder strikte condities beschikbaar zijn voor een specifieke groep Bescherming is vereist door wet-/regelgeving De VU Amsterdam is verplicht zich te melden bij de Functionaris Gegevensbescherming en aan de personen als er op ongepaste wijze toegang tot de gegevens wordt verkregen Het verlies aan vertrouwelijkheid van de gegevens zou een desastreuze negatieve gevolgen hebben voor onderzoeksdoelen, veiligheid, budget of reputatie Het gebruik van deze gegevens kan een aanzienlijke impact hebben op een iemands leven	Transcripties van interviews met (LHBTI) vluchtelingen die praten over hun thuisland Video-interviews met kinderen die praten over misbruik Open tekstreacties (bijv. feedback in dagboekvorm) van patiënten met mentale of fysieke aandoeningen/ beperkingen Genetische gegevens van kwetsbare personen die wijzen op een risico op ziekte of aandoeningen Onderzoek waarbij inkomensgegevens van burgers gebruikt worden

Bijlage 2: Basismaatregelen

Beschikbaarheid

Klasse	Maatregelen			
	Toegang	**Verwerking**	**Opslag**	**Datacommunicatie**
Laag	n.v.t.	n.v.t.	Dagelijkse data back-up Opslag op 2 locaties	n.v.t.
Midden	Back-ups van de data kunnen in het geval van een calamiteit zo snel mogelijk teruggezet worden.	Er is minimaal een acceptatie- en een productieomgeving	Dagelijkse data back-up, snapshots per uur Opslag op 2 locaties en 1 offline variant Er is een actueel overzicht van welke data opgeslagen zijn en in hoeverre deze data volledig zijn	n.v.t.
Hoog	Back-ups van de data kunnen in het geval van een calamiteit zo snel mogelijk teruggezet worden.	Er is minimaal een acceptatie- en een productieomgeving	Dagelijkse data back-up, snapshots per uur Opslag: 3 kopieën, en 2 verschillende media types, waarvan 2 off-site Er is een actueel overzicht van welke data opgeslagen zijn en in hoeverre deze data volledig zijn Er is een uitwijkvoorziening ingericht voor de fysieke opslag van de data Bestanden worden op een beperkt toegankelijke plek op het netwerk bewaard.	n.v.t.

Integriteit

Klasse	Maatregelen			
	Toegang	**Verwerking**	**Opslag**	**Datacommunicatie**
Laag	Minimaal gebruikersnaam en wachtwoord expliciet toegekend aan eigen medewerkers (conform het wachtwoordbeleid) Inlogactiviteit wordt gelogd	n.v.t.	n.v.t.	n.v.t.
Midden	*Maatregelen hierboven, aangevuld met:* Multi-factor authenticatie bij toegang van buiten de VU-infrastructuur Inlogactiviteit, als ook wijziging van persoonsgegevens of andere kritische gegevens wordt gelogd Er is actieve monitoring op de logging	Versiebeheer van documenten	Toegang tot de informatie waar data is opgeslagen dient te worden beveiligd met een sterk wachtwoord	n.v.t.
Hoog	*Maatregelen hierboven, aangevuld met:* Sterke vormen van multi-factor authenticatie (denk hierbij aan Time-based One-Time Password, Tokens, biometrisch, etc.), onafhankelijk van locatie Inlogactiviteit, als ook inzage en wijziging van persoonsgegevens en andere kritische gegevens wordt gelogd Er is actieve monitoring op de logging, als ook periodieke reviews van de logs op de integriteit Toegang tot het informatiesysteem waarmee data wordt verwerkt dient te worden beheerst met behulp van een authenticatieproces met gebruikmaking van smart cards of biometrische lezers	*Maatregelen hierboven, aangevuld met:* Validatie op input van data Training voor sleutelgebruikers	Toegang tot de informatie waar data is opgeslagen dient te worden beveiligd met een sterk wachtwoord multi-factor authenticatie	Gebruik van handtekening bij data-transport

Vertrouwelijkheid

Klasse	Maatregelen			
	Toegang	**Verwerking**	**Opslag**	**Datacommunicatie**
Laag	n.v.t.	n.v.t.	n.v.t.	n.v.t.
Midden	Alleen geautoriseerde personen mogen toegang hebben Voor externe toegang: Lijst van geautoriseerde personen waarin de eigenaar van de data de namen specificeert of functies of personen die toegangsrechten hebben op de betreffende data Multi-factor authenticatie is optioneel, maar aanbevolen bij diensten die standaard de functionaliteit meeleveren	Data-encryptie tijdens transport: encryptie op extern netwek Het scherm waarop data worden getoond dient automatisch op slot te gaan na 15 (?) minuten van inactiviteit	Het informatiesysteem waarmee data worden verwerkt mag alleen in ruimten worden geplaatst met fysieke toegangsbeveiliging Data zijn versleuteld (optioneel maar aanbevolen)	Wanneer bestanden worden uitgewisseld via diensten als SFTP, instant messaging, enz., dan dienen ze te worden beveiligd met een wachtwoord Encryptie op extern netwerk De verzender dient zorgvuldig de ontvanger te controleren
Hoog	Voor alle toegang: Lijst van geautoriseerde personen Beperkt tot strikt noodzakelijke gebruikers en rechten Multi-factor authenticatie	*Maatregelen hierboven, aangevuld met*: Alleen de dataeigenaar mag data verwijderen Gegevens moeten worden verwijderd alleen door middel van een algoritme dat een veilige vernietiging garandeert Gebruikers dienen uit het informatiesysteem waarmee data worden verwerkt te loggen indien zij tijdelijk of blijvend de werkplek hebben verlaten	*Maatregelen hierboven, aangevuld met*: Data zijn versleuteld bij opslag: sleutel is in bezit bij de data geautoriseerde personen (eventueel ook voor beheerders) Alleen personen met een autorisatie voor deze data mogen toegang hebben tot het deel van het informatie-systeem waar het document is opgeslagen	*Maatregelen hierboven, aangevuld met*: E-mail dient te worden versleuteld indien deze naar buiten de organisatie wordt verzonden
Zeer hoog	Voor alle toegang: Lijst van geautoriseerde personen Beperkt tot strikt noodzakelijke gebruikers en rechten Sterke vormen van multi-factor authenticatie (denk hierbij aan Time-based One-Time Password, tokens, biometrisch, etc.), onafhankelijk van locatie	*Zie maatregelen hierboven.*	*Maatregelen hierboven, aangevuld met:* Data zijn versleuteld bij opslag: sleutel is alleen in bezit bij de data geautoriseerde personen Beheerders hebben alleen toegang tot de encrypted data en hebben geen sleutel De data mogen alleen worden opgeslagen op servers die worden beheerd en/of ondersteund door de VU en geschikt bevonden zijn voor betreffende toepassing (o.a. Scistor, Yoda, Research Drive)	*Maatregelen hierboven, aangevuld met:* Encryptie op zowel intern en extern netwerk De data mogen niet uitgewisseld worden via diensten zoals FTP, instant messaging, etc.

Bijlage 3: Classificatie RDM informatiesystemen[7]

Classificatie	Yoda	Research Drive	Scistor 1.0	DataverseNL	OSF
Laag	ja	ja	ja	ja	ja
Midden	ja	ja	ja	ja	nee
Hoog	ja	ja*	ja*	nee	nee
Zeer hoog	ja*	nee	nee	nee	nee

Tabel 3.1. Overzicht van door de VU aanbevolen en ondersteunde data-management oplossingen en respectievelijke classificaties op de as Vertrouwelijkheid. (* Alleen mogelijk met aanvullende maatregelen). Naast deze aanbevolen data-management oplossingen kunnen onderzoekers ook gebruikmaken van standaard IT-infrastructuur die niet specifiek bedoeld is voor verwerking van onderzoeksdata. Dit zijn onder andere OneDrive, Microsoft Teams en GoogleDrive. Gegevens over deze applicaties zijn te vinden in de Data Storagefinder.

When to use

Yoda	Research Drive	Scistor	DataverseNL	OSF
Storage of large volumes of data that don't need to be accessed frequently for processing/analysis	Storage of large volumes of data that need to be regularly accessed for processing & analysis	Storage of very large volumes of data that need to be accessed regularly for processing/analysis	For publishing data openly or with restrictions for at least 10 years	For publishing data, documentation and protocols and sharing privately with others
Creation of structured metadata to describe your research and the associated datasets	Similar uses to SURFdrive, but ensures that data storage is linked to a project rather than an individual	Data can be accessed directly from SciStor without copying to a local drive prior to processing & analysis	For data that are not confidential (anymore)	Possibility to connect external storage (Research Drive, Dataverse) to OSF.
Offers read-only storage of (a copy of) your raw data	Has a desktop sync client for easy management of locally copied data	Best option for high-performance computing due to direct communication with servers	Generates DOIs and allows adding metadata	For creating and/ or publishing preregistrations
Allows data to be shared with internal and external (non-VU) collaborators but only at dataset-level	Allows for finegrained access management at the folder and subfolder level	Best option for connecting directly to lab devices		For data that are not confidential
Provides archiving for data after a research project is published	Allows for easy collaboration and sharing with external collaborators	There are many additional options for versioning, backup, security, etc.		Includes research project management functionalities

[7] Bij deze tabellen moet worden aangetekend dat deze de huidige stand van zaken weergeven en dat dit aan verandering onderhevig is, mede vanwege doorontwikkeling van de applicaties zelf en de beveiligingsmaatregelen die eraan gekoppeld kunnen worden.

When not to use

Yoda	Research Drive	Scistor	DataverseNL	OSF
Not efficient for the storage of large volumes of data that need to be regularly accessed for processing/analysis	Requires encryption for higher risk data	Cannot be used for collaboration with external (non-VU) users	Not for personal or otherwise confidential data if no consent has been given to publish data (also with restrictions)	Threshold for free storage is lower than Dataverse or Yoda
Difficult and slow to access data directly on the YODA disk; data will likely need to be copied locally prior to data processing/analysis	Not possible to interact directly with the Research Drive disk; requires syncing of data locally before processing/ analysis	Does not offer free storage up to 500 GB like YODA and Research Drive do		Not for personal or otherwise confidential data if no consent has been given to publish data (also with restrictions)
Lacks a native desktop sync client for easy management of local copies of data	Does not offer structured metadata documentation to help make data FAIR	May not be appropriate for storage of higher risk data		Does not create DOIs automatically, you need to actively select it
Does not allow for access management of subfolders; everyone in your YODA group folder has access to all subfolders therein	Does not provide locking or vault options to prevent raw data from being modified or to serve as an archive	Does not offer structured metadata or documentation to help make data FAIR		
		If using SciStor from home, connectivity ends up similar to YODA; advantages are only present when connecting on campus		

Research Data Management

rdm.vu.nl

VU VRIJE UNIVERSITEIT AMSTERDAM