

ARX 3.8.0

	Data anonymization tool	
	Systeemvereisten:	Win, OS-X of Linux (32 of 64 bits)
	Ontwikkeld door:	Community driven
Apache License, Version 2.0	Contactpersoon:	Kristof.verslype@smals.be

Functionaliteiten

ARX is een uitgebreide tool om gevoelige persoonsgegevens zo goed als mogelijk te anonimiseren. De tool richt zich op *microdata*, dus data met records op het niveau van het individu. Naarmate er echter meer informatie is per individu, zal ook het anonimiseren zonder de data nutteloos te maken, moeilijker worden.

In een eerste fase wordt data geïmporteerd; Na wat voorbereidend werk kan het resultaat geanonimiseerd worden. Dat houdt in dat identificerende attributen zoals rijksregisternummers weggelaten worden en data andere gegeneraliseerd worden. Dat laatste kan fijnmazig op het niveau van het individuele record; Om tot een bepaald niveau van bescherming te komen kan het voor sommige records het geboortejaar behouden blijven, terwijl voor andere records slechts een geboortecategorie weerhouden wordt. Een dergelijke flexibiliteit is nodig om het optimale evenwicht te vinden tussen enerzijds het minimaliseren van het privacy risico en anderzijds het minimaliseren van het informatieverlies (en dus nutsverlies). De analysetools, die deel uitmaken van ARX, helpen de gebruiker om op gefundeerde wijze dit evenwicht te vinden. Ten slotte kan de geanonimiseerde dataset geëxporteerd worden naar één van de ondersteunde formaten.

De algoritmes die ARX ondersteunen, waaronder *k-anonymity*, *l-diversity* en *t-closeness* zijn gebaseerd op wetenschappelijk & gepubliceerd onderzoek. Verder kan ofwel gebruik gemaakt worden van de grafische user interface, ofwel van de API.

Conclusies & Aanbevelingen

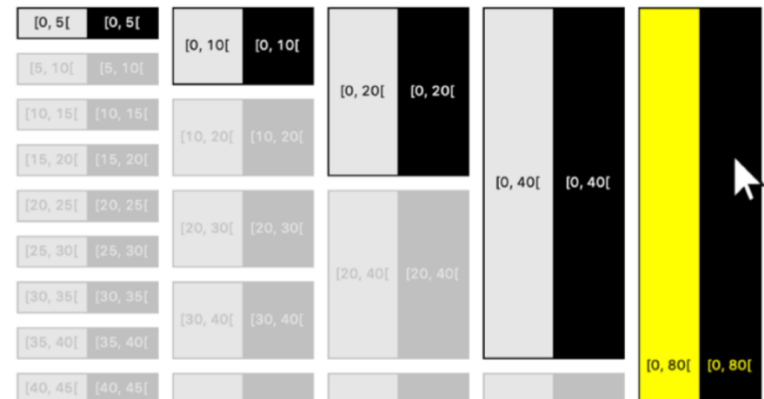
ARX is een degelijke en zeer volledige tool voor het anonimiseren van persoonsgegevens. Daarbij moet benadrukt worden dat het enkel kan werken indien de individuele records een beperkt aantal attributen bevatten. Bij grote hoeveelheden persoonsgegevens per record gaan we richting big data en wordt anonimiseren sowieso onmogelijk zonder de data de facto onbruikbaar te maken.

Het gebruik van de tool is niet steeds even evident zonder voorafgaandelijke kennis. De leercurve kan dus als vrij steil ervaren worden.

Testen & Resultaten

De testen werden uitgevoerd tijdens een hand-on workshop georganiseerd door de KU Leuven. Er werd vertrokken van een fictieve dataset bestaande uit meer dan 30 000 records, die telkens 9 attributen bevatten, waaronder geslacht, leeftijd, opleiding en burgerlijke status. Tijdens de test werd uitsluitend gebruik gemaakt van de GUI (zie <https://github.com/arx-deidentifier/arx/blob/master/data/adult.csv>).

Na het importeren, moeten de data voorbereid worden. Dit houdt onder meer het opstellen van generalisatieregels in. Stel bijvoorbeeld dat de exacte leeftijd te specifiek is, hoe mag het systeem generaliseren? In dat geval is het aan de gebruiker om aan te geven dat bijvoorbeeld op het eerste niveau categorieën per 5 jaar genomen moeten worden en op het tweede niveau categorieën per 10 jaar (zie afbeelding). Verder moet ook het privacy model gekozen worden. In ons geval kozen we voor *k-anonymity*.



Daarna stelt ARX heel wat anonimisatiepistes voor. In de ene piste kan bijvoorbeeld het geslacht wat meer gegeneraliseerd zijn, terwijl dat in een andere piste de opleiding is. Voor elke piste kun je een erg gedetailleerd inzicht verwerven in zowel de risico's als het informatieverlies. Verschillende types aanvallen worden daarbij in rekening gebracht. De onderstaande figuur geeft een samenvatting van de risico's bij een *prosecutor attack*, wat wil zeggen dat de aanvaller op zoek is naar informatie over één specifieke, vooraf gekozen burger.



Gebruiksvoorwaarden & Budget

ARX is open source software ontwikkeld en onderhouden door vrijwilligers, vaak uit de academische wereld. Er is geen betalende enterprise versie of betalende support.