

Five W's of Open Data.





TU Delft library met Open Data Expert and [Data Champion, Anneke Zuiderwijk](#), to learn about the 'who, what, where, when and why' of Open Data.

Anneke Zuiderwijk is a [Data Champion](#) working within the [Faculty of Technology, Policy and Management](#). As an Assistant Professor in the Department of [Engineering, Systems and Services](#), she has dedicated her academic career to investigating the complex rationale behind open data. In Zuiderwijk's unique case, data sharing isn't just a research consideration, rather it's her research domain.

Zuiderwijk believes that great value can be created through open data sharing and use and that it can be revolutionary for scientific advancement. Her mission is to develop theory for designing infrastructures and institutional arrangements that incentivize open data sharing so that it eventually becomes the standard rather than the exception.

In order to realise her mission, her academic research and responsibilities as a Data Champion address the following '5W's of open data':

1. Who are the actors involved?
2. What are their motivations to share data? And, what can infrastructure do to motivate data sharing?
3. Where can data be shared?
4. When should they prepare to share data?
5. Why consider data sharing?

About Anneke

Before joining TU Delft in 2011, Zuiderwijk worked for the [Dutch Ministry of Justice and Security](#) where she gleaned insight into the real world of open judicial data. "The Ministry were collecting data on crime statistics and were faced with many challenges of sharing such sensitive information with the wider community."

Motivated to mitigate the challenges associated with sharing and using government data, Zuiderwijk began her PhD under the supervision of [Professor Marijn Janssen](#) at TU Delft. Her [doctoral research](#) involved the design of a socio-technical infrastructure that enhanced the coordination of open government data use. By creating a metadata model with interaction and data quality mechanisms, Zuiderwijk aimed to stimulate interaction between data providers, data users and policy makers thereby creating a feedback loop to better inform policy about government data use.

After obtaining her PhD with distinction, Zuiderwijk assumed a postdoctoral researcher position working on the [VRE4EIC](#) project, funded by the European Union's 3-year Horizon 2020 research and innovation programme. 'VRE4EIC' stands for 'Virtual Research Environment to Empower multidisciplinary research communities and accelerate Innovation and Collaboration'. Essentially, this online platform draws professionals from various scientific domains to exchange data, resources and knowledge within a multidisciplinary data environment. Data can be made open at different levels within the platform meaning that there is control over data sharing and data can be protected if necessary.

Zuiderwijk emphasises the importance of embracing diversity and fostering inclusion for scientific progression. "Many research environments focus on discrete scientific disciplines with limited convergence, yet, as scientific problems are related across all disciplines, it's of tremendous value to bring disciplines together to solve problems collectively. After all, societal problems must be studied from multiple perspectives."

1. Who are the actors involved?

The field of open data involves a number of actors. "Researchers, governments, companies, citizens, journalists, librarians and archivists are all concerned with open data to some extent," informs Zuiderwijk. A key objective of her research is to understand the needs of these different actors and how to motivate them to openly share their data and use the data of others. "In order to build a sociotechnical infrastructure that facilitates effective data sharing across multiple disciplines, it's necessary to understand the perspectives, behaviours and motivations of each actor involved."

2. What are their motivations to share data? And, what can infrastructure do to motivate data sharing?

She explains that their motivations are diverse and discipline-specific. "Researchers may be motivated to openly share data if it means greater visibility and more citations to boost their career development." However, Zuiderwijk admits, "There are still obstacles that prevent researchers from sharing data, such as a lack of the required skills."

Whilst the majority of obstacles cannot be mitigated completely, Zuiderwijk explains that implementing suitable infrastructure and related institutional arrangements can help. "It's important to improve the ease of use of data infrastructures, of which open data portals are one element, and to provide institutional training for using such portals so that researchers acquire the necessary skills and become motivated to share data."

Data sharing is more common in some academic disciplines than others and there are obvious disparities across them. Her recent publication, [‘Sharing and re-using open data: A case study of motivations in astrophysics’](#), explores the complex interaction of factors influencing open data sharing and use in astrophysics, a single scientific discipline where data is already extensively shared and reused. Her case study demonstrates that benefits associated with data sharing involve intrinsic motivations, such as more reproducible science and faster rates of scientific advancement. Zuiderwijk hopes that insights gained from her study of astrophysics can be transferred to disciplines where data sharing is less common.

She discusses the motivations of other actors. “Whilst governmental organisations are typically motivated to share data for public value, it’s more difficult for corporate organisations to do so.” She continues, “Commercial enterprises are often restricted by proprietary interests, trade secrets, and concerns over legal ramifications of privacy and security breaches.” Zuiderwijk is ambitious to help overcome the barriers of data sharing and use within the private sector and build trust through the creation of mechanisms for sharing open business data. This is an objective she wishes to explore in the near future.

3. Where can data be shared?

Data is typically shared via online platforms. Zuiderwijk uses the [4TU.ResearchData](#) repository to openly publish her data. She believes that one of her most important contributions as a Data Champion is motivating others to follow her practice. “I actively encourage my students to publish their data on the 4TU repository.” She adds, “Working transparently gives them more credibility as early career researchers. If a reviewer can see that their underlying data is openly accessible it instils confidence that the findings presented within their manuscript are robust and reliable.”

4. When should they prepare to share data?

As soon as Zuiderwijk’s students begin their research, she educates them on the principles of Open Science and proper research data management. She ensures that they each receive relevant training to openly publish their data upon completion of their research project. “From the point of data collection, students are trained on how to cleanse and curate their data so that it’s understandable to others when they openly share it. Formatting data correctly saves time and effort when it comes to uploading data to the repository.”

Zuiderwijk coordinates with faculty [Data Steward, Nicolas Dintzner](#), to help students create their data management plans prior to conducting research. Wherever necessary, she also mandates that students apply to the research ethics committee so that they understand the regulations and requirements of collecting and sharing confidential data.

“Reluctance to share confidential data often arises because researchers believe that data anonymisation costs extra time and effort.” She argues against this belief, using her own research methodology as an example. “When I systematically analyse personal data, such as interview transcripts, I assign codes to the data to describe the content. This means the codebook; an Excel spreadsheet containing analysed data, is already anonymised. With no extra labour, the coded version can be openly published on the repository to benefit the wider community.”

5. Why consider data sharing?

Despite important considerations for confidentiality, there’s no doubt that the arguments for sharing data are powerful. Individuals have an opportunity to advance their scientific knowledge, and improve their visibility and credibility, which in turn raises their professional profile. They have more scope to build international, multidisciplinary connections and foster a collaborative community wherein resources can be shared for social and economic gain.

Whilst the benefits are obvious, Zuiderwijk admits that the culture change towards open data is slow. “It’s a long-term transition. It took me several years to appreciate why I should share my data,” she says. “I’m fortunate that in my field of research I get to experience the benefits of data sharing and reuse. Unfortunately, actors in other disciplines rarely get to witness the advantages.”

Passionate about educating others about the importance of data sharing, Zuiderwijk has instructed the ProfEd course on [Open Data Governance: From Policy and Use](#) and the Massive Open Online Course (MOOC), [Open Science: Sharing Your Research with the World](#), together with TU Delft Library instructors, [Michiel de Jong](#) and [Nicole Will](#). Alongside the leader of TU Delft’s [Knowledge Centre Open Data](#) (Kenniscentrum Open Data), [Bastiaan van Loenen](#), she will lecture during the upcoming [MOOC on Open Government](#) organised by [Marijn Janssen](#), which takes place in September this year.

Citation

This showcase is a selection from the following publication:

Clare, Connie. (2019). The Real World of Research Data [Book]. Zenodo. <http://doi.org/10.5281/zenodo.3584373>.

