

Kristin Kaltenhäuser (2022): Destabilising Data in Nordic Asylum Decision-making. In: Proceedings of the 20th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies - Doctoral Colloquium, Reports of the European Society for Socially Embedded Technologies (ISSN), DOI: 10.48340/ecscw2022_dc06

Copyright 2022 held by Authors, DOI: 10.18420/ecscw2022_dc06

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, contact the Authors.

Destabilising Data in Nordic Asylum Decision-making

Kristin Kaltenhäuser

Department of Computer Science, University of Copenhagen

Contact Author: krka@di.ku.dk

Abstract. Asylum decision-making is a complex collaborative work domain. The probability of receiving asylum for individuals from the same country of origin varies significantly across states. Research in this area has been conducted in different disciplines, such as legal studies, social sciences and data science. It remains fragmented and applies discrete methodologies that are rarely integrated. I will combine qualitative work/domain expertise and participatory dialogue with computational decision modelling to answer two questions: 1) What factors shape the production of national asylum decisions? and 2) Why do asylum outcomes across similar cases differ so much from one another? I aim to map a set of interdisciplinary methodological and conceptual tools for engaging asylum decision-making data that can lead to the discovery of possible missing data and "counter datasets." In a preliminary study, I investigated gender-related categories of an open dataset of 9.075 Danish asylum case summaries using data science methods and applying an archival perspective. The analytical insights will be used to facilitate the grounded sensemaking of data together with different groups of practitioners.

1 Studying Asylum Decision-making by Destabilising Data

Asylum decision-making is a highly complex collaborative work domain. The probability of receiving asylum for individuals from the same country of origin varies significantly across states. For example, in 2018 Somali applicants had an 8% chance of receiving asylum in the first instance in Denmark, against 34% in Norway, and 48% in Sweden. Research in this area has been approached from different disciplines, such as legal studies, social sciences and data science, yet it remains fragmented. Discrete methodologies are applied that are rarely integrated (Noll et al. (2007)) and fail to describe and support the complex collaborative work domain sufficiently.

Artificial Intelligence for example has been used in various countries to develop computational models to uncover the influence of extra-legal factors on decision outcomes (e.g. gender factors, adjudicators' level of experience), yet have often fallen short to be sufficiently representative to adequately support the decision process (Schmidt and Bannon (2013)). Instead of developing yet another machine learning algorithm, my interest lies in destabilising the data that is used as a basis for computational decision models and finding novel ways to engage with it. Data can be destabilised by employing interdisciplinary methods and exposing its localised and biased nature (D'Ignazio and Klein (2020)), as well as confronting the data with the perspectives of its subjects.

Inspired by recent work (Menendez-Blanco et al. (2017), Ismail and Kumar (2018), Møller et al. (2021), Aragon et al. (2022)), I aim to combine qualitative work/domain expertise and participatory dialogue with computational decision modelling to answer the following two questions: 1) What factors shape the production of national asylum decisions? and 2) Why do asylum outcomes across similar cases differ so much from one another? I plan to map a set of interdisciplinary methodological and conceptual tools for engaging asylum decision-making data that can lead to the discovery of possible missing data and "counter datasets." (D'Ignazio and Klein (2020)) and ultimately to a more nuanced understanding of this complex collaborative work domain.

My methodological approach combines quantitative computational analysis with qualitative methods. Computational analysis is used to assess the relative importance of different features of asylum cases and develop predictive models of asylum outcomes. Qualitative analysis of legal documents and participatory methods enable grounded and critical sense-making of data together with practitioners and policy makers.

Star and Strauss (2004) and later Bechmann (2019) and Møller et al. (2021) write about giving voice to silenced perspectives and suggest grounded sensemaking of data together with the data subjects to correct the underlying power imbalance which data subjects face. I aim to build on this approach and confront the data about asylum decision-making with its subjects and data workers.

The following sections outline an initial analysis of Danish asylum case data from an archival perspective (Thylstrup et al. (2021)). Drawing from data science methods, qualitative analysis of legal documents and taking an archival perspective, with this preliminary study I aimed to study, how gender-related categories are presented in the Danish asylum case summaries, as a way of destabilising asylum data. I first investigated the construction of the category of gender-related persecution by the United Nations Refugee Agency (UNHCR) through a collection of topics, such as homosexuality and female genital mutilation. Based on these topics, I analysed an open dataset of 9.075 asylum case summaries, handled by the Danish Refugee Appeals Board and conducted a topic analysis to determine how the formal category of gender-related persecution was transformed in a national context. This study resulted in analytical insights that can be used as material to engage with stakeholders and to study the practices of asylum decision-making in Denmark.

2 An Archival Perspective as an Entry Point to Studying Asylum Decision-Making

Increasingly, data in asylum decision-making is saved to databases (both national and international) that come to form an archive, a long standing interest of CSCW Ackerman et al. (2013). Following Foucault's notion of power, an archive can be an empirical manifestation functioning as a tool for data collection and production, but also an analytical concept or a lens (Foucault (1972)). Power from this perspective is 'generative' and archives are distinctly uncertain spaces of knowledge production (Thylstrup (2022)). Taking an archival perspective thus also means paying close attention to power structures that shape archives D'Ignazio and Klein (2020) and thus become scrutable (Thylstrup (2022)).

I deconstruct the empirical gender-related categories archived as part of the collaboration across the Danish national authorities and the Danish Refugee Appeals Board. Deconstructing categories is an example of destabilising data in the sense that it brings to the fore topics that are omitted and articulated in the empirical categories. While *practices* of categorization and classification are a long-term interest of CSCW (Suchman (1993), Bowker and Star (1999), Møller and Bjørn (2011), Boyd and Crawford (2012), Pine and Liboiron (2015)), an archival perspective seeks to increase sensitivity towards understandings context and power relationships structuring datasets. My interest is to establish the empirical gender-related categories as a link between the categorization practices that take place on a national level and the international political context that shape it.

3 Gender-related Persecution in International Asylum Law

While international law - notably the 1951 Refugee Convention - establishes a common definition of who counts as a refugee, states adapt and transform this category through both their national law and decision-making. States therefore operate different institutional and procedural frameworks for asylum decision-making. This process results in loosely coordinated but still diverging practices across states.

The 1951 UN Refugee Convention describes a refugee as:

owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country. (Article 1 A (2) Refugee Convention, UNHCR (1951))

Unlike some later human rights conventions, gender or sexual orientation is not a category explicitly listed in this definition. Since the 1980s, however, national asylum authorities have gradually come to recognise that gender may form the basis of an asylum claim, for instance where women face repercussions for transgressing social mores or sexual minorities face widespread persecution in the societies in which they live.

In 2002, the UNHCR issued a guideline on gender-related persecution (UNHCR (2002), which is illuminating in terms of elaborating upon and qualifying the types of cases which belong within the category of gender-related persecution. As such, the UNHCR guideline may be read as constructing a category for seeking asylum. Topics that are included in that category are homosexuality, female genital mutilation etc. In the ensuing analysis, I systematically extract these topics, query the case summaries on this basis and analyse their occurrence in the empirical categories applied by the Danish Refugee Appeals Board.

4 Using NLP Methods to Interrogate Categories

The object of this analysis is the publicly available dataset by the Danish Refugee Appeals Board (<https://fln.dk/praksis>). The data consists of summaries of asylum decisions by the Refugee Appeals Board in Denmark between 2004 and 2021. The Refugee Appeals Board is the second institution to assess applications for asylum in Denmark. Only cases rejected by the Danish Immigration Service, the first instance in the process, are automatically referred to the Refugee Appeals Board.

I perform the analysis of the topic of gender-related persecution in steps:

	Gender-related persecution	LGBT	Sexual conditions	Marital conditions	Other gender-related persecution	Total
Homosexuality	9%	89%	12%	1%	1%	269
Rape	17%	12%	21%	30%	19%	262
Female genital mutilation	8%	0%	2%	17%	65%	177
Forced marriage	7%	0%	5%	41%	14%	143
Human trafficking	6%	1%	1%	1%	11%	48
Forced abortion	1%	1%	3%	10%	2%	41
Bisexuality	0%	15%	1%	0%	0%	39
Forced prostitution	4%	2%	2%	1%	5%	35
Transgenderism	0%	2%	0%	1%	0%	6
Forced sterilisation	1%	0%	0%	0%	1%	6
Transvestism	0%	0%	0%	0%	0%	0
Total	303	211	366	474	175	

Figure 1. This Figure shows the 11 topics extracted from the UNHCR guideline on gender-related persecution (rows) and the five empirical gender-related categories in the public dataset of the Danish Refugee Appeals Board (columns). It illustrates the prevalence of different topics in the categories. The percentages indicate the proportion of overall cases in the category that are tagged with the topic. One case can be tagged with several topics and some cases are not tagged with one of the extracted topics at all. Thus, the percentages of the categories (in the columns) don't add up to 100%. The darker the cell shading, the bigger the proportion of cases in that particular category that mention a particular topic.

1. I conducted initial exploratory analysis of asylum motives in the data and the UNHCR definition of gender-related persecution. I then selected the categories in the Danish dataset that match the definition given by the UNHCR. These are the empirical categories:

- gender-related persecution
- LGBT
- sexual conditions
- marital conditions
- other gender-related persecution

Following the UNHCR guideline, claims based on sexual orientation contain a gender element, and include topics related to the umbrella term LGBT. I therefore included the categories "LGBT" and "Sexual conditions" in my analysis.

2. I then systematically extracted topics from the UNHCR guideline on gender-related persecution, that are explicitly stated as claims related to gender. Table 4 provides a list of the extracted topics in the first column.

3. Finally, I conducted a topic analysis by querying the case summaries for words related to the extracted topics, see Figure 1. Table I lists the queried terms for each topic. I employed manual stemming of the query words to reduce words to their basic form or stem. This technique is often applied in NLP to account for different forms of words in a text. When selecting

keywords to represent the topics, I aimed to stay as close to the wording of the topic as possible. Each case that contains any of these terms, is counted as one occurrence. That means a case can contain several topics. I then calculated the percentage composition of each category to illustrate which topics constitute a category.

Topic	Queried terms
Homosexuality	homoseks, lesbisk
Human trafficking	menneskehandel, traffick
Female genital mutilation	omskåret, omskæring
Forced marriage	tvangsgift
Forced prostitution	prostitution
Rape	voldt
Transgenderism	transkøn
Forced abortion	abort
Forced sterilisation	sterilis
Bisexuality	biseks
Transvestism	transvest

Table I. List of topics extracted from the 2002 UN guideline on gender-related persecution, as well as the terms that represent a topic in a case.

5 Destabilising Data on Gender for Understanding Practice

An overall theme for my research is the destabilisation of data. The central premise of this research is to not take data as a given, but to trace its origin and the processes of interpretation that shaped it. Figure 1 showcases one example of how I work with destabilising of data through topic analysis. The topic analysis is reorganising knowledge and in this sense links to a long-term debate of CSCW on knowledge production (Ackerman et al. (2013)). I illustrate how data science techniques can be useful for reorganising and probing data for omitted and articulated topics in asylum case summaries. Figure 1 shows how categories are composed of topics. Some topics are dispersed over several categories (e.g. human trafficking, rape or forced prostitution), while others are mainly located in one category (e.g. homosexuality). Some of these topics are more articulated than others by almost being turned into their own homogeneous category, as is the case with the category of LGBT, of which 89% of cases are about homosexuality. Other topics, such as female genital mutilation are omitted and therefore rendered invisible by being placed in a residual category. Residual categories cover the cases that do not fit into "pure" categories (Bowker and Star (1999)). One thus expects a

wide scope of topics to be included in that category. However, the empirical residual category in the Danish asylum case data *other gender-related persecution* is dominated by the topic of female genital mutilation. The destabilising of data in this case is exemplified by showing how topics are transformed into categories or not as a result of the work practices underlying asylum decision-making.

I plan to use the initial insights into the data of Danish asylum case data as a basis for participatory sensemaking of the dataset together with various stakeholders of the asylum decision-making process in Denmark. The tables shown in my research description can be seen as a starting point for the design of data artefacts that can be used to facilitate a participatory dialogue.

An obvious next step is to trace the effects of omitted topics, for example, female genital mutilation: Is the recognition rate of asylum lower for this particular group? What does it mean to connect this insight to practice by taking a participatory approach? How can a participation-inspired NLP unfold in a complex, collaborative work domain such as asylum, where we as scholars act as intermediaries between interests of stakeholders with differential power?

References

- Ackerman, M. S., J. Dachtera, V. Pipek, and V. Wulf (2013): 'Sharing Knowledge and Expertise: The CSCW View of Knowledge Management'. *Computer Supported Cooperative Work*, vol. 22, no. 4, pp. 531–573.
- Aragon, C., S. Guha, M. Kogan, M. Muller, and G. Neff (2022): *Human-Centered Data Science - An Introduction*. MIT Press.
- Bechmann, A. (2019): 'Data as Humans: Representation, Accountability, and Equality in Big Data'. *Human Rights in the Age of Platforms*.
- Bowker, G. C. and S. L. Star (1999): *Sorting Things Out: Classification and Its Consequences*, Inside Technology. MIT Press.
- Boyd, D. and K. Crawford (2012): 'Critical Questions for Big Data'. *Information, Communication & Society*, vol. 15, no. 5, pp. 662–679.
- Byrne, R. and T. Gammeltoft-Hansen (2020): 'International Refugee Law between Scholarship and Practice'. *International Journal of Refugee Law*, vol. 32, no. 2, pp. 181–199.
- Bødker, S., C. Dindler, O. S. Iversen, and R. C. Smith (2021): *Participatory Design*, Vol. 14. Morgan & Claypool Publishers.
- D'Ignazio, C. and L. F. Klein (2020): *Data Feminism*, Strong Ideas. MIT Press.
- Feinberg, M. (2017): 'A Design Perspective on Data'. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA, p. 2952–2963, ACM Press.
- Foucault, M. (1972): *The archaeology of knowledge*. Tavistock Publications.
- Gitelman, L. (ed.) (2013): *"Raw Data" Is an Oxymoron*, Infrastructures. MIT Press.
- Ismail, A. and N. Kumar (2018): 'Engaging Solidarity in Data Collection Practices for Community Health'. *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, pp. 1–24.

- Kensing, F. and J. Blomberg (1998): 'Participatory Design: Issues and Concerns'. *Computer Supported Cooperative Work*, vol. 7, pp. 167–185.
- Kitchin, R. (2014): *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE.
- Latour, B. (1999): *Pandora's Hope: Essays on the Reality of Science Studies*. Harvard University Press, 1st edition edition.
- Menendez-Blanco, M., P. Bjorn, and A. De Angeli (2017): 'Fostering Cooperative Activism through Critical Design'. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. New York, NY, USA, p. 618–629, ACM Press.
- Møller, N. L. H., G. Fitzpatrick, and C. A. Le Dantec (2019): 'Assembling the Case: Citizens' Strategies for Exercising Authority and Personal Autonomy in Social Welfare'. *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. GROUP.
- Møller, N. H. and P. Bjørn (2011): 'Layers in Sorting Practices: Sorting out Patients with Potential Cancer'. *Computer Supported Cooperative Work*, vol. 20, no. 3, pp. 123–153.
- Møller, N. H., G. Neff, J. G. Simonsen, J. C. Villumsen, and P. Bjørn (2021): 'Can Workplace Tracking Ever Empower? Collective Sensemaking for the Responsible Use of Sensor Data at Work'. *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, pp. 219:1–219:21.
- Noll, G., D. E. Anker, and M. D. Muller (2007): 'Explaining Credibility Assessment in the Asylum Procedure'. *International Journal of Refugee Law*, vol. 19, no. 3, pp. 599–602.
- Pine, K. H. and M. Liboiron (2015): 'The Politics of Measurement and Action'. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. pp. 3147–3156, ACM Press.
- Randall, D., R. Harper, and M. Rouncefield (2007): *Fieldwork for Design: Theory and Practice*. Springer.
- Schmidt, K. and L. Bannon (2013): 'Constructing CSCW: The First Quarter Century'. *Computer Supported Cooperative Work*, vol. 22, no. 4–6, pp. 345–372.
- Seidelin, C., Y. Dittrich, and E. Grönvall (2020): 'Foregrounding data in co-design – An exploration of how data may become an object of design'. *International Journal of Human-Computer Studies*, vol. 143, pp. 102505.
- Star, S. L. and A. Strauss (2004): 'Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work'. *Computer Supported Cooperative Work (CSCW)*, vol. 8, pp. 9–30.
- Suchman, L. (1993): 'Do Categories Have Politics? The Language/Action Perspective Reconsidered'. In: *Proceedings of the Third Conference on European Conference on Computer-Supported Cooperative Work*. USA, p. 1–14, Kluwer Academic Publishers.
- Thylstrup, N. B. (2022): 'The ethics and politics of data sets: deleting traces and encountering remains'. *Media, Culture & Society*.
- Thylstrup, N. B., D. Agostinho, A. Ring, C. D'Ignazio, and K. Veel (eds.) (2021): *Uncertain Archives: Critical Keywords for Big Data*. MIT Press.
- UNHCR (1951): *The Refugee Convention, 1951*. Geneva, Switzerland.
- UNHCR (2002): *Guidelines On International Protection: Gender-Related Persecution within the context of Article 1A(2) of the 1951 Convention and/or its 1967 Protocol relating to the Status of Refugees*. Geneva, Switzerland.