Are we able to extract useful and quality data from online sources to paint a picture of reuse of textiles in Norway?

Mona Eide Onstad, Statistics Norway, Mona.Onstad@ssb.no

Manju Chaudhary, Statistics Norway, Manju.Chaudhary@ssb.no

Gisle Berge, Statistics Norway, Gisle.Berge@ssb.no

Abstract

With an increasing focus on reuse and circular economy in the EU and UN, there is a demand for statistics on reuse. Through the Directive of the European Parliament and of the Council (EU) 2018/851, Norway is required to report on the reuse of textiles. To meet this growing demand, we began investigating how we can use online sources to find data on textile reuse in Norway.

Our perception is that a considerable portion of the market of reuse is online, through web pages such as Finn.no, Tise, Fretex and Facebook Marketplace. Any analysis on reuse would be incomplete without the inclusion of online platforms. Thus, a lot of data should be readily available for us for statistical purposes. The objective of this paper is twofold. First, investigate the accessibility of the data – is it as readily available as it appears? We investigated both the quick access solution of web scraping and data received through contact with online platforms. The second objective of the paper is to assess the quality of the data acquired through these methods. Specifically, whether it is usable and complies with our needs for official statistics.

Through a sensitivity analysis this study provides provisional estimations for reuse of textile sold or given away on online platforms in Norway to be somewhere in the range of 770 and 2 845 tonnes (between 1-3.6 percent of the total consumer consumption of new clothes and home textiles).

As more data is becoming available through new, online sources, it is important to research whether it can lessen the workload that goes into surveys and other methods for establishing new statistics and can be a good quality data source. In this paper we share the experiences that we have gained, dipping our toes in this new area.

Keywords: reuse, online data sources, digital platforms, circular economy, sensitivity analysis, web scraping

1. Introduction

With the impending climate crisis, society is at a crossroads. There has been a political shift, especially in the EU/Europe, from an economy that is largely based on selling new items that turn into waste, to aspiring towards a more circular economy. With their circular economy action plan (European Commision, 2020), the EU seek to increase the life span of resources and decrease waste.

As of today, Statistics Norway only have data on one aspect of circular economy – waste. We measure how many resources are removed from the economy as waste, but not how much is reused. As more countries adapt strategies of circular economy, it is critical to keeps tabs on its development and measure its success.

The field of reuse is vast and expands through all parts of the economy. Traditional methods of obtaining these statistics, such surveys, is quite demanding when it comes to resources, both in term of time and money. A lot of reuse happens online, especially when it comes to households purchasing second hand textiles and furniture. We are curious to see whether we can access these resources quickly and with ease from online sources, and whether the data has good enough quality to be used for official statistics.

The paper starts by looking into web scraping, highlighting its potential, but also its disadvantages. We move on to looking at data that has been provided voluntarily from major online platforms in Norway. As the webpages are quite distinct from each other, so is the data. We describe how we combine these different datasets and how we analyze them. Lastly, we look at some drawbacks of these methods and give recommendations for further work.

2. Web scraping

The internet has become a major market in terms of re-use of textiles and clothes. Using the web through scripted and automated web scraping was one of our first method to try out in order to collect and study online data.

2.1 Limitations to the method

Although web scraping may provide detailed data relatively quickly, and poses few limitations, there are a few disadvantages that make us not recommend this method for general statistical purposes, which are:

- Websites commonly don't want automatic web scraping activities and bots to operate on their websites/servers (take up resources on their servers, want to protect their own data etc.). For that reason, they may also implement measures to prevent or limit web scraping activities on their site. The legality surrounding web scraping is also somewhat in a "grey zone", but normally one gets an indication of what the different websites prefer of activity while staying on their website from the robots.txt files, which most websites add to their site right after their domain name (e.g. https://www.website.no/robots.txt).

- Websites commonly change in content or technical format over time, and this may cause existing code and scripts for data extraction to break upon running later and requires rewriting the code.

- Data derived from web scraping activities is usually not followed by any good metadata description. One can see the data, but it is not always obvious what you look at behind the "technical window".

- In terms of re-use of textiles and clothes, readily available data for web scraping also doesn't always mean clothes being sold and changing hands in a reuse market. It reflects more the potential of reuse and a static window of the situation of what is available at a certain time.

2.2 Advantages of web scraping

Despite these limitations, web scraping can still provide valuable insight at an early phase of the process and an overview of the data's potential as a data source for statistical purposes. There is a lot of available data out on the internet – some easy and some difficult to access – and sometimes also very detailed data, like geographical coordinates and information.

2.3 Example data

The following data is derived from web scraping a website specializing in reuse of shoes and clothes items (the name of the site is intentionally concealed here), using web scraping techniques in Python (specifically, the libraries Requests and BeautifulSoup).

The data scraped consists of material from two different time periods, one month around Christmas season in December/January 2021 (scraping date: 16.01.2021), and the other around Easter in April 2022 (scraping date: 29.04.2022).

What also is uncertain here is the effect of sold items throughout the times series, since advertised items will unavoidably reduce the further it is away from the scraping dates. Thus, in the following two bar plots, a downward trend is to be expected the further away one is from these two dates.
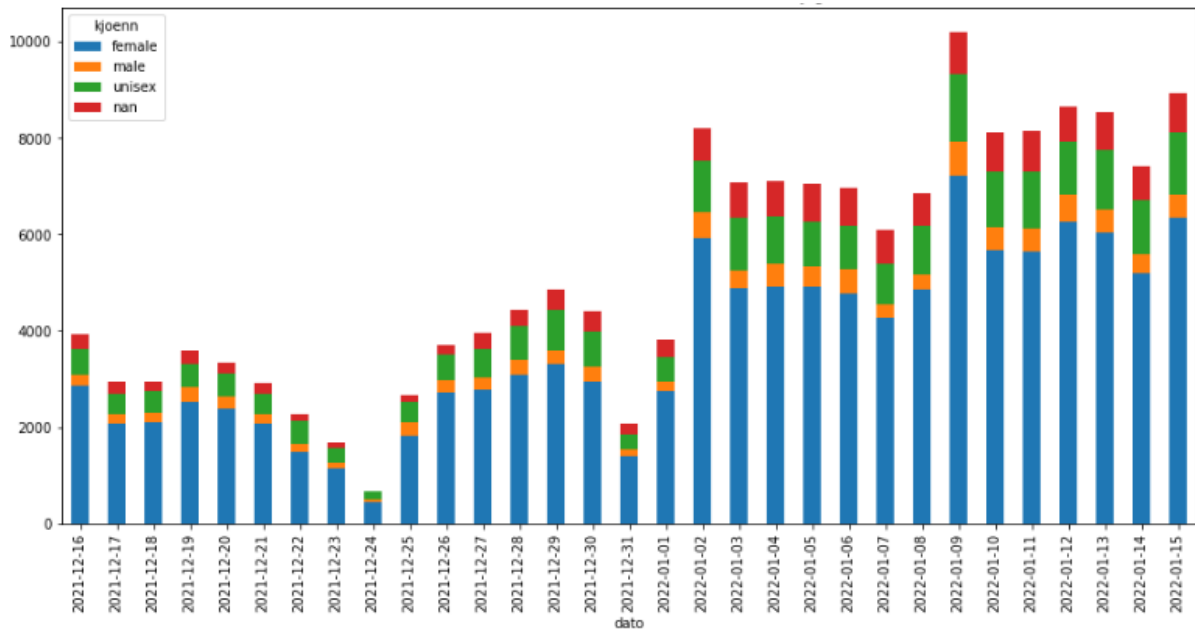

Figure 1: Number of advertisements - clothes and shoes - based on webscraping a random website in December/January 2021.
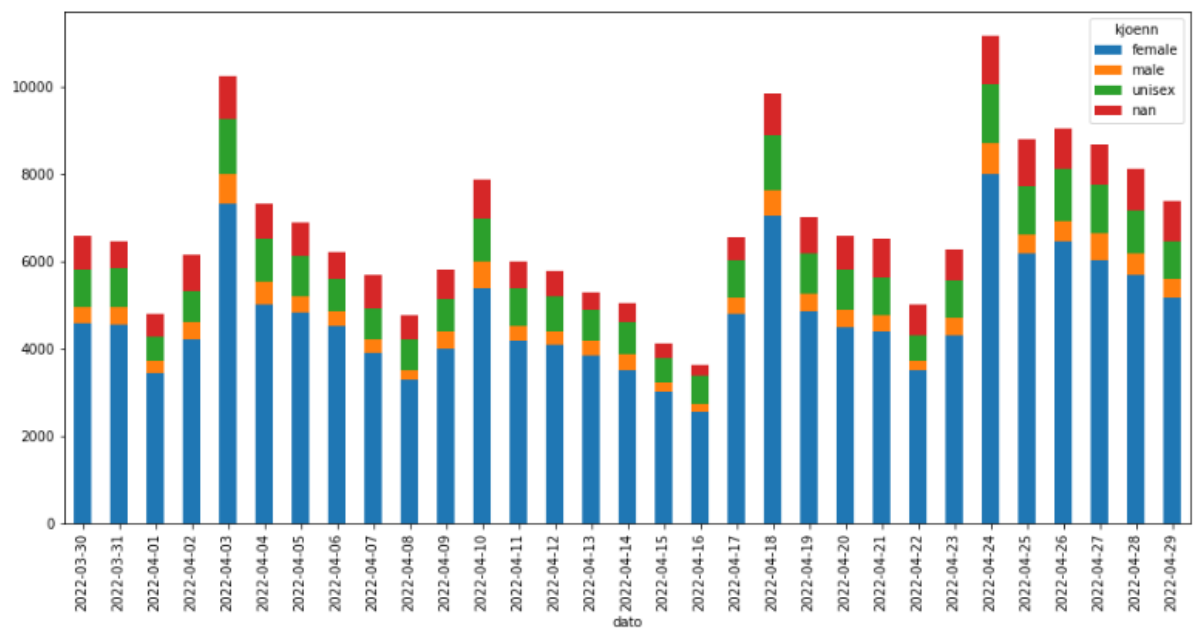

Figure 2: Number of advertisements - clothes and shoes - based on webscraping a random website in April 2022.

There are some apparent cyclical effects over time in the data, and especially the Christmas time series before and after New Year, then there appears to be a steep increase of number of advertisements out on this website. Seems unwanted Christmas gifts going straight to second-hand sale.

It is also not uncommon to come across geographical information in the scraped material, like coordinates or postal address, and thus a geographical dimension may be possible to generate. The red dots on the map below illustrate all the 165 000 advertisements available around April 2022, and they can readily be plotted on a map without any further preparation.

Figure 3: Geographical locations based on a web scraping a random website in April 2021.

As to be expected, the re-use market in shoes and clothes is dominant in larger cities and urban areas.

3. Gathering data by contacting online platforms

Due to the issues associated with web scraping, we decided to go forward by contacting the owners of the data platforms and try to get first-hand data from them.

Compared to web scraping, the process was more tedious and granted poorer, less detailed data and results. It was a process filled with emails and online meetings, before finally receiving the data in Excel files by email. We were able to gather data from the two major platforms, Finn.no and Tise, as well as Fretex's the online store. We were not able to get in touch with Facebook to ask for data from Facebook Marketplace.

In this section of the article, we will describe the data that we were able to gather, detail how we combined the data, do some analysis and finally discuss its limitations.

3.1 Describing the data

The datasets from the three different sources came in distinct formats, much like the outline of their web page.

3.1.1 Finn.no

The data received by Finn.no did not have detailed data per item of clothing or tonnes of textiles sold. They provided us with the number of advertisement posts that had been put up on their webpage, with total numbers for 2021. We were also provided with a percentage of posts marked as sold in total for textiles, not based on the different categories. The data was divided into the categories that they use on their web page. For each category these is the number of posts, as well as the average price. There were five main categories that have textiles, 9 subcategories and 49 secondary subcategories in total. For example, the main category clothes, cosmetics and accessories has the subcategory women's clothing, and 12 secondary subcategories as follows:

Wedding dresses
Dresses
Hats, scarfs, and mittens
Jackets
National costume (bunad)
Outerwear
Shirts
Skirts

| **Sweaters** |
| **T-shirts** |
| **Tops** |
| **Trousers** |
| **Underwear** |
| **Other** |

The data was further divided into counties and municipalities. Through this voluntary exchange of data, they did not provide us with details like city of origin for these advertisements. We would have preferred to have data on each advertisement instead of aggregated data in categories.

3.1.2 Tise

Unlike Finn.no, the data from Tise is in terms of sold piece of clothing. They were only able to provide us with data for the last three months of 2021. It is divided according to the following categories:

| **Date** |
| **City** |
| **Size** |
| **Gender** |
| **Category** |
| **Asking price** |

Gender refers to the gender of the seller, and not necessarily of the clothing item. Asking price is the price stated in the advertisement, while the price finally agreed between the seller and buyer might be different. The categories of clothing are even more specific than those of Finn.no. There is a total of 114 subcategories for textiles. An example of categories are "wearables", "clothes" and "sweaters" as the final subcategory.

3.1.3 Fretex

Fretex is major when it comes to selling second-hand items in physical stores, but somewhat smaller when it comes to online sales. Their online store provided us with data on sold items from their internet sales. This was the least detailed data, with no details on categories of textiles. We were provided with the number of sold items each month of 2021 and the average price.

Turnover including VTA
Number of goods sold
Average price including VTA
Average price excluding VTA

3.2 Combining the data

There are several steps that must be taken in order to combine the datasets and conduct some analysis. In this section we go through how we adjust the data before delving into the analysis in section 4.

3.2.1 Cleaning up the data

The first step is to "clean up" the data so that the data sets have comparable data. To do this we had to specify our definition of textiles. We decided that it should include shoes, bags, bed sheets and even sleeping bags. By employing a specific definition, we were able to remove all the variables that did not comply with it.

3.2.2 Inflating the data from Tise

The data from Tise is only for the last 3 months of 2021. It must be inflated to a whole year so that we can combine it with the data from Finn.no and Fretex. We applied a simple method of multiplying the 3 months data with 4 so that it is 12 months in total. The weakness of this approach is that there could be seasonal differences across the year that we are not able to account for.

3.2.3 Determining the percentage sold

The data from Finn.no told us how many advertisements there were in 2021 selling or giving away used textiles. It does not tell us whether it was sold or given away. On an aggregate level, Finn has provided us with the percentage of posts that were marked

as sold in 2021 for textiles. This will always be an underestimate because not everyone cares to mark their advertisement as sold – some just delete them.

3.2.4 Converting number of advertisements to kilograms

Number of advertisements in a year is not a very useful unit for analysis. We want to investigate the weight of reused textiles and need to convert the data into more familiar measurements. The data from Finn.no and Tise is quite detailed in terms of type of textile. We used this information in combination with theoretical factors of average weights of different types of textiles from a shipping company (Shippingstorm.com, 2022). Then we were able find the total weight of reused textiles and clothes bought online by multiplying the number of items from Tise and Finn.no with average weight factors from the shipping company.

The data from Tise contains only one piece of clothing per advertisement, unless otherwise specified. In the Finn.no data, however, some advertisements could contain more than one textile. This is sometimes the case with children's clothing and could lead to an underestimation of weight of textiles. We have not been able to go into detail with that in our analysis here.

The data from Fretex did not provide us with many details, only the number of items sold. To find an approximate weight per advertisement, we assume that what is sold online with Fretex is like what is sold in Finn.no and Tise. By making that assumption we found an average weight per item from Finn.no and Tise and multiplied number with the data from Fretex.

3.2.5 Controlling for duplicates

It is possible that a person posts an item on several online platforms to increase their chances of a sale. It is also possible to mark an item as sold across all platforms, even though it might have been sold on Tise and not Finn, or vice versa.

According to Tise, the data we received from them only includes sold items. Finn also gave us the percentage of clothes sold in 2021. If they are certain that the items were sold on their web page, duplicates should not be a problem. Due to the nature of online platforms, however, it seems unlikely that they can be certain. With more detailed data we could potentially check for duplicate advertisements across platforms posted at around the same time and location. However, we do not possess this data now, so we are unable to control for duplicates.

3.2.6 Combining the data

The last step was to create one combined dataset from the three online platforms that we have. We had to use all three so that the data is protected and cannot be identified back to one company. Fretex provided the data with the fewest details in terms of category of textiles. We therefore had to aggregate the rest of the data, accordingly, only showing a total sum. The data from Finn had annual totals, and not months like Tise and Fretex. Thus, we also aggregated those two for a whole year. We then had the total amounts of reused textiles in tonnes sold online on these platforms, as well as the average and total price and were ready to conduct our analysis.

3.3 Analysis of data and results

The dataset we have certainly has some limitations, but by using a sensitivity analysis we are able to get an idea of the range of the amount of online textile reused in Norway. With this approach we will consider some of the uncertainties that remain in our dataset and analysis.

The sensitivity analysis involves calculating the amount of textile reused for different scenarios – meaning different values of the uncertain parameters of the model.

There are three major dimensions of uncertainties we had to take into consideration in our analysis, because we do not have complete knowledge of them now:

- We do not know the percentage of Finn data that is sold out of total advertisements, only that it is at the very minimum 40 per cent. It may also be the case that an advertisement can be sold simply by deletion, not informing Finn.no when it happens. Thus, this needs to be taken into consideration, and we will make scenarios also including 50, 60 and 70 per cent sold.

- The overlap of advertisements between Finn.no and Tise is largely unknown. In order to take into consideration that the same advertisement can be put out on both platforms, we need to make some estimations of how big the overlap can be. We employ a range of possible values of overlap between the two to be explored in the sensitivity analysis: 0, 15, 30, 45 and 60 per cent.

- Facebook is another big "black box" in the equation. So far we have not been able to acquire any data or contact with them, but we know this is likely a large contributor in the second-hand market of reused textiles and clothes. Uff is another participant we know of, smaller than Facebook, but selling clothes on

Instagram and not included in our dataset. Thus, in order to include sales in missing data sources like Facebook, we have, based on knowledge from the dataset from Tise and Finn.no come up with a range of possible scenarios. These scenarios differ depending on what type of uncertainty analysis we want to calculate (number of advertisements, amount of textiles sold in tonnes and percent, and amount of textiles sold in unit million NOK). Possible value range is presented in the last four columns in table below.

Type of uncertainty analysis	Percentage sold Finn.no				Percentage overlap Finn.no vs. Tise.com					Amount sold on Facebook etc.			
Number of advertisements	40	50	60	70	0	15	30	45	60	1 million	1,5 million	2 million	2,5 million
Textiles sold (tonnes)	40	50	60	70	0	15	30	45	60	100 tonnes	500 tonnes	1 000 tonnes	1 500 tonnes
Textiles sold (per cent)	40	50	60	70	0	15	30	45	60	100 tonnes	500 tonnes	1 000 tonnes	1 500 tonnes
Textiles sold (million NOK)	40	50	60	70	0	15	30	45	60	300 mill. NOK	600 mill. NOK	900 mill. NOK	1200 mill. NOK

Table 2 The setup of the sensitivity analyses showing all tested values applied in the different scenarios.

We have four different uncertainty analysis in order to describe each of the following dimensions of the data:

- Number of advertisements sold
- Textiles sold (unit: tonnes)
- Textiles sold (unit: percent)
- Textile sold (unit: million NOK)

Within each of these four uncertainty-analysis we end up with 80 different scenarios for the amount of second-hand textiles sold in Norway (equal to the multiplication of 4 values for *Percentage sold Finn.no*, 5 for *Percentage overlap Finn.no and Tise.com* and 4 from *Amount sold on Facebook etc.*).

These 80 values are presented in the following sections below, one uncertainty analysis after another.

3.3.1 Uncertainty analysis: Tonnes of reused textiles sold online

The following table shows the results in green to red color of the 80 different scenarios for tonnes of reused textiles sold online in 2021. They vary from the smallest estimate being 770 tonnes textile reuse from online platforms to 2 845 as the highest estimate.

Percentage sold Finn.no	Percentage overlap Finn.no vs Tise.com	Amount sold on Facebook etc.			
		100 tonnes	500 tonnes	1 000 tonnes	1 500 tonnes
40	0	1 130	1 530	2 030	2 530
	15	1 040	1 440	1 940	2 440
	30	950	1 350	1 850	2 350
	45	860	1 260	1 760	2 260
	60	770	1 170	1 670	2 170
50	0	1 235	1 635	2 135	2 635
	15	1 145	1 545	2 045	2 545
	30	1 055	1 455	1 955	2 455
	45	965	1 365	1 865	2 365
	60	875	1 275	1 775	2 275
60	0	1 340	1 740	2 240	2 740
	15	1 250	1 650	2 150	2 650
	30	1 160	1 560	2 060	2 560
	45	1 070	1 470	1 970	2 470
	60	980	1 380	1 880	2 380
70	0	1 445	1 845	2 345	2 845
	15	1 355	1 755	2 255	2 755
	30	1 265	1 665	2 165	2 665
	45	1 175	1 575	2 075	2 575
	60	1 085	1 485	1 985	2 485

Table 2: Sensitivity analysis of the amount of textiles sold on online platforms in Norway, 2021 (unit: tonnes)

3.3.2 Uncertainty analysis: Percentage of reused textiles sold online

Related and as a continuation of the result above, we wanted to find out roughly how much of the reused amounts are compared the total market of new textiles in Norway.

According to Norsus (2020) Norwegian consumers buy around 80 000 tonnes (15 kg per person) of clothes and home textiles a year. Using that number as a basis for how much is "out there", our findings show between 1 to 3.6 per cent are being sold and bought online in the reuse market (with less than 5 per cent would be a good "rule of thumb").

Percentage sold Finn.no	Percentage overlap Finn.no vs Tise.com	Amount sold on Facebook etc.			
		100 tonnes	500 tonnes	1 000 tonnes	1 500 tonnes
40	0	1,4 %	1,9 %	2,5 %	3,2 %
	15	1,3 %	1,8 %	2,4 %	3,0 %
	30	1,2 %	1,7 %	2,3 %	2,9 %
	45	1,1 %	1,6 %	2,2 %	2,8 %
	60	1,0 %	1,5 %	2,1 %	2,7 %
50	0	1,5 %	2,0 %	2,7 %	3,3 %
	15	1,4 %	1,9 %	2,6 %	3,2 %
	30	1,3 %	1,8 %	2,4 %	3,1 %
	45	1,2 %	1,7 %	2,3 %	3,0 %
	60	1,1 %	1,6 %	2,2 %	2,8 %
60	0	1,7 %	2,2 %	2,8 %	3,4 %
	15	1,6 %	2,1 %	2,7 %	3,3 %
	30	1,4 %	1,9 %	2,6 %	3,2 %
	45	1,3 %	1,8 %	2,5 %	3,1 %
	60	1,2 %	1,7 %	2,3 %	3,0 %
70	0	1,8 %	2,3 %	2,9 %	3,6 %
	15	1,7 %	2,2 %	2,8 %	3,4 %
	30	1,6 %	2,1 %	2,7 %	3,3 %
	45	1,5 %	2,0 %	2,6 %	3,2 %
	60	1,4 %	1,9 %	2,5 %	3,1 %

Table 3: Amount of textiles sold on online platforms in Norway, 2021, sensitivity analysis (unit: percent of total sale, with total sale estimated to 80 000 tonnes, see NORSUS (2020))

3.3.3 Uncertainty analysis: Number of advertisements of reused textiles sold online

Similarly, the total number of advertisements for reused textiles sold in 2021 can be illustrated. The range varies from 2,5 million to 5,6 million advertisements, depending on choice of scenario variables.

Percentage sold Finn.no	Percentage overlap Finn.no vs Tise.com	Number of advertisements sold on Facebook etc.			
		1 million	1,5 million	2 million	2,5 million
40	0	3 509 177	4 009 177	4 509 177	5 009 177
	15	3 264 702	3 764 702	4 264 702	4 764 702
	30	3 020 227	3 520 227	4 020 227	4 520 227
	45	2 775 751	3 275 751	3 775 751	4 275 751
	60	2 531 276	3 031 276	3 531 276	4 031 276
50	0	3 723 941	4 223 941	4 723 941	5 223 941
	15	3 479 465	3 979 465	4 479 465	4 979 465
	30	3 234 990	3 734 990	4 234 990	4 734 990
	45	2 990 514	3 490 514	3 990 514	4 490 514
	60	2 746 039	3 246 039	3 746 039	4 246 039
60	0	3 938 704	4 438 704	4 938 704	5 438 704
	15	3 694 228	4 194 228	4 694 228	5 194 228
	30	3 449 753	3 949 753	4 449 753	4 949 753
	45	3 205 277	3 705 277	4 205 277	4 705 277
	60	2 960 802	3 460 802	3 960 802	4 460 802
70	0	4 153 467	4 653 467	5 153 467	5 653 467
	15	3 908 991	4 408 991	4 908 991	5 408 991
	30	3 664 516	4 164 516	4 664 516	5 164 516
	45	3 420 041	3 920 041	4 420 041	4 920 041
	60	3 175 565	3 675 565	4 175 565	4 675 565

Table 4: Number of sales on online platforms in Norway, 2021, sensitivity analysis (unit: number of sales)

3.3.4 Uncertainty analysis: Size of online second-hand market on textiles

The prices also made us analyze the size of the online reuse market. The data from Fretex constitute the selling price, while Finn.no and Tise have information of the asking price in the advertisement. It is entirely possible that there have been price negotiations between buyer and seller, but it is believed for our purpose that this constitutes a good enough indicator of the selling price.

The table below shows according to our sensitivity analysis that the market of online second-hand textiles in Norway could potentially vary from 1 100 to 2 800 million NOK.

Percentage sold Finn.no	Percentage overlap Finn.no vs Tise.com	Amount sold on Facebook etc.			
		300 mill NOK	600 mill. NOK	900 mill NOK	1 200 mill NOK
40	0	1 594	1 894	2 194	2 494
	15	1 465	1 765	2 065	2 365
	30	1 335	1 635	1 935	2 235
	45	1 206	1 506	1 806	2 106
	60	1 077	1 377	1 677	1 977
50	0	1 699	1 999	2 299	2 599
	15	1 570	1 870	2 170	2 470
	30	1 441	1 741	2 041	2 341
	45	1 311	1 611	1 911	2 211
	60	1 182	1 482	1 782	2 082
60	0	1 805	2 105	2 405	2 705
	15	1 675	1 975	2 275	2 575
	30	1 546	1 846	2 146	2 446
	45	1 417	1 717	2 017	2 317
	60	1 288	1 588	1 888	2 188
70	0	1 910	2 210	2 510	2 810
	15	1 781	2 081	2 381	2 681
	30	1 652	1 952	2 252	2 552
	45	1 522	1 822	2 122	2 422
	60	1 393	1 693	1 993	2 293

Table 5: Sensitivity analysis of the amount of textiles sold on online platforms in Norway, 2021 (unit: million NOK)

To put this in perspective the total Norwegian online market is estimated by NETS (2021) to be around 156.7 billion NOK in 2020, and out of that, clothes alone constitute 14,5 billion NOK. Our reuse data from the sensitivity analysis above will then level to somewhere between 7 to 19 per cent of the new clothes online market in financial terms.

4. Limitations to our data

Our assumption is that a big share of textile reuse happens through online platforms in Norway, but we cannot ignore that there are also other physical platforms where the ownership for such textile products changes. For example, physical second-hand stores which sell second-hand textiles, private gifts or donations, flea markets etc.

Physical second-hand stores could be owned by both private companies and charity organizations. Many non-profit charity organizations collect, sort and reuse textiles in Norway as well as export to other countries. These types of sales and reuse of textiles have not been included in the scope of our analysis, as we have only focused on data collections from digital platforms.

We have included the data from three out of four main online platforms for second-hand sale of textiles. The three are Finn.no, Tise and Fretex, and the one missing is Facebook Marketplace. As mentioned earlier we have not been able to get in contact with Facebook, and for that we do miss out on one major data source. Still based on information from other key online platforms – Finn.no, Tise and Fretex – we have tried to make some qualified guesses based on what we can make out of the data from them. As a temporary solution in this project, with Facebook missing, we have tried to accommodate this by including them in the sensitivity analysis, using different values for possible sales for Facebook in the calculations. Since levels presented in the sensitivity analysis, only give rough levels – and a range – thus more work needs to be done into more detailed data, precise results and filling in gaps of uncertainty that remains in the sensitivity analysis. Therefore, it is noteworthy that large unreachable digital company could make such a big complication in data collection in this digital age.

Possible duplicates of sales and advertisements across online platforms poses another challenge in our calculation. From the available datasets we could not properly identify the duplicates as we do not have a common identification variable. Also in this project, the datasets we received had aggregations, as some datasets were detailed advertisements data, while others were aggregated. For statistical purposes, it is preferable to get detailed data on advertisement level for all the datasets, not aggregated, in order to fully explore and set up a possible duplication removal procedure. As we did not have detailed data for all datasets, the possible overlap between Finn.no and Tise was taken into account through the sensitive analysis, being one of the dimensions there.

However, one conceivable solution, if we had all non-aggregated datasets, could be to use geographical information of addresses or coordinates, perhaps also with a time restriction, in order to remove duplicates in the datasets. But for now, it is an untried option in this project due to the aggregated structure of some of the input data. It is potential to overcome this challenge by requesting the non-aggregated data from data providers according to required specifications. This exercise could be time consuming for first time but in a scenario of regular delivery, automation of such procedures is very much possible.

We received data for sold items from all three volunteer data providers. To generate data for sold items one needs to mark the advertisement as sold, but some people just delete the entry without marking it with a successful sale or sold item. Therefore, in some cases sold items might get underestimated. A possible solution could be technical function on websites which does not allow one to delete the entry without mentioning if the item in question is sold or not or if sold on other platforms. So, we could reduce the underestimation error.

A challenge in this study we have met is that the number of sold items per online advertisement can be more than one. It causes difficulties in calculating both average weight and price of items. In such cases conversion factors cannot be used appropriately to convert the number of textile articles to weight unit. With a more detailed data set it should be possible to identify when a post contains several items of textile. There is an uncertainty of underestimation due to this fact, but we have not in this project been able to adjust for this without any additional information.

We have also noticed a few news releases on cases of false advertisement at some of these online platforms, which are presumably negligible but as an error source we cannot ignore on a later stage and should be considered during data collection and data editing. But for this project we assume that digital platforms keep this problem at bay.

5. Conclusive remarks

With more and more digitalization of required services in our daily life, we generate data in abundance on daily basis. This data can be used to develop and produce statistics, and sale of second-hand textile items is no different. These data are easily available and can be retrieved from the digital platforms in question.

In this paper we set out to answer two questions; whether data on reuse was easily accessible from digital platforms, and whether it was of high enough quality to use for development of official statistics. We believe we have managed to answer both questions and got a good overview and insight of textile reuse in Norway. There is no doubt that access to quality data improves timeliness, reduces costs of statistics development and lowers respondent's burden when these data are replacing a survey or other time-consuming data collection methods.

We have explored web scraping for this study, which is a quick way of gathering a lot of useful data but with copyright restrictions, relatively extensive data collection and little reliable metadata we decided not to go further down this road. Thus, we do not recommend the approach until and unless, one has taken care of these hindrances in advance.

When it comes to gathering data directly from online platforms, we encountered several obstacles already mentioned. We believe that process could have been more effective by using the statistics act of Norway (Statistikkloven, 2019). Which provides improved access to all types of data, including privately held data. We could then have expected better quality, data delivered according to our requirements and broader representation. Thus, many of the limitations in our study can be dealt with.

This study gives a good overview of sale of second-hand textiles on digital platforms, and it can be a good base for further studies and in developing a statistic on reuse of textiles in Norway. Though to have a complete overview of second-hand sale of textiles, one should not forget to consider sales in physical stores, private donations, flea markets etc. Further to develop an official statistic, we must comply with quality requirements like those of European statistics, i.e., the European Statistics Code of Practice (Eurostat, 2022).

6. References

European commission 2020: *New circular economy action plan.*

[EUR-Lex - 52020DC0098 - EN - EUR-Lex (europa.eu)](EUR-Lex)

Shippingstorm.com 2022: *Approximate weights of goods.* Available from: https://shippingstorm.com/en/list-of-weight/

NORSUS 2020: *Kartlegging av brukte tekstiler og tekstilavfall i Norge.* Available from: https://norsus.no/publikasjon/kartlegging-av-brukte-tekstiler-og-tekstilavfall-i-norge/.

NETS 2021: *Norsk E-handel 2020.* Available from: https://info.nets.no/norsk-ehandelsrapport-2020

Statistikkloven 2019: Lov om statistikk og Statistisk sentralbyrå (statistikkloven). (Eng. The statistics law). Available from: https://lovdata.no/dokument/LTI/lov/2019-06-21-32

Eurostat 2022: *Safeguarding quality by 16 principles*. Available from:

https://ec.europa.eu/eurostat/web/quality/european-quality-standards/european-statistics-code-of-practice