



Transport  
Innovation  
Gender  
Observatory

# Deliverable 4.4

## Workshop on gender and diversity in digital analysis

[www.tinngo.eu](http://www.tinngo.eu)



This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 824349

**Disclaimer:** This document reflects only the author's view. Responsibility for the information and views expressed therein lies entirely with the authors. The Innovation and Networks Executive Agency (INEA) and the European Commission are not responsible for any use that may be made of the information it contains.

## Deliverable fact sheet

Document name	Workshop on Gender and Diversity in Digital Analysis	
Responsible partner	UCPH	
Deliverable number	4.4	
Due date of deliverable	December 2019	
Actual submission date	Resubmission date December 2020	
Version	Version 2	
Authors	Hilda Rømer Christensen Michala Hvidt Breengaard Stine Pedersen	
Reviewers	Andree Woodcock	
Work package no.	WP4	
Work package title	Workshop on Gender and Diversity in Digital Analysis	
Work package leader	UCPH	
Work package participants	Reports from TINNGO hubs: Baltic, Romanian, Greek, German/French, UK, Nordic, and mixed team.	
Dissemination level (please select one)		
CO	Confidential, only for members of the consortium including the Commission Services	
PU	Public	X
PP	Restricted to other programme participants, including the Commission Services	
RE	Restricted to a group specified by the consortium including the Commission Services	
Nature of the deliverable (please select one)		
R	Report	X
D	Demonstrator	
W	Websites, patents filing, etc	
O	Other	

## Abstract

This report is based on the TInnGO workshop on gender and diversity in digital analysis, convened at the University of Copenhagen, November 19<sup>th</sup>, 2019. The aim of the workshop was to educate all TInnGO project partners in how to apply a gender and diversity focus to the analysis of digital, visual material. Digital analysis presents a new mode of knowledge production which enables comparative analysis of various gender diversity and transport discourses. The overall objectives of the workshop were to widen the methodological scope of the project and to introduce all TInnGO partners, using visual media archives as data material. The digital workshop also provided a kickoff event for the targeted digital analysis, reported as a TInnGO deliverable D. 4.6. ([https://www.tinn.go.eu/wp-content/uploads/2020/11/TInnGO\\_D4.6\\_Media-Analysis-Report.pdf](https://www.tinn.go.eu/wp-content/uploads/2020/11/TInnGO_D4.6_Media-Analysis-Report.pdf))

## Contents

Abstract	3
Contents	4
Summary:	5
1. Introduction	7
2. Conduct of the workshop	9
2.1 TInnGO Digital analysis workshop: Program and introduction by the team from Aalborg University: .....	11
2.2 Program.....	12
2.3 Summary of the Workshop.....	13
3. TInnGO Digital analysis workshop	14
3.1 UK Hub: Gender and diversity in Smart Mobility .....	14
3.2 Baltic Hub: Lithuania-UK comparison on future mobility .....	16
3.3 Romanian Hub: Romania-Denmark comparison, Women in Management ....	16
3.4 UCPH and IR Hub: Romania-Denmark comparison, employment and technology .....	17
4. General Assessments and Potentials	18
4.1 The Romanian Hub.....	18
4.2 German and French Hub.....	18
4.3. Mixed team ( Inter + UCop).....	18
4.4 Greek Hub .....	19
5. Summary: Digital Analysis as a learning process	20
5.1 Widening and focusing: Search topics and findings .....	20
5.2 Technical knowledge and supervision .....	20
Appendix	22
1. TInnGO digital analysis	22
2. Additional guidelines	22
3. Copyright and Open Access	23
4. Gender and diversity in Digital Analysis	23

## Summary:

The aim of this report is to present the TInnGO digital analysis as a possible research method for studies of transport, gender and diversity. The method was presented at a digital workshop in Copenhagen in 2019. The workshop was attended by 32 TInnGO participants, who experimented with and explored the usefulness of connecting gender, diversity and transport in analysis of visual data in their respective countries.

The workshop was conducted by the TInnGO partner at University of Copenhagen in co-operation with experienced researchers from University of Aalborg in Denmark.

### **Aims of the digital workshop:**

- To widen the methodological knowledge of TInnGO partners, and to provide a kickoff workshop for the later, targeted digital analysis.
- To provide methods of digital analysis as a supplement to the more traditional methods used in research projects in general, and TInnGO in particular, such as quantitative and qualitative data methods e.g. statistics analysis, surveys and interviews.
- To focus on gender smart mobility transposed into visual representations of autonomous cars and transport as a gendered labour market.

### **Methods and Findings:**

- A digital tool was developed beforehand and presented to the TInnGO partners: the Technical Google Trends was applied to find relevant national / and comparative search queries, combined with the Google Image Scarper, the Clarifai recognition / tagging data service and the visualizing Gephi tool.
- The TInnGO partners provided evidence of the tool's usefulness for assessing gender imbalances in visual representations of gender and technical intersections (e.g. the overrepresentation of technical devices and general under representation of humans).
- As for the regional differences, they turned out to be quite weak, which indicate that smart transport and autonomous cars represent an emerging discourse.

Given existing differences in mobility systems and practices, this conclusion was not expected, and might be qualified through the collection of more data.

**Perspectives and outreach:**

- The digital tools used by TinnGO partners were assessed as having significant shortcomings, but having the potential to shed light on relevant TinnGO themes e.g. the gendered representations in smart transport.
- Digital analysis cannot stand alone, but can be seen as a useful and stimulating supplement to other research methods, such as the use of conventional statistics and interviews and mappings related to various transport modalities and practices.
- The TinnGO digital tools have subsequently been applied as educational material and exercise material in the MA: Social Data Science at University of Copenhagen.

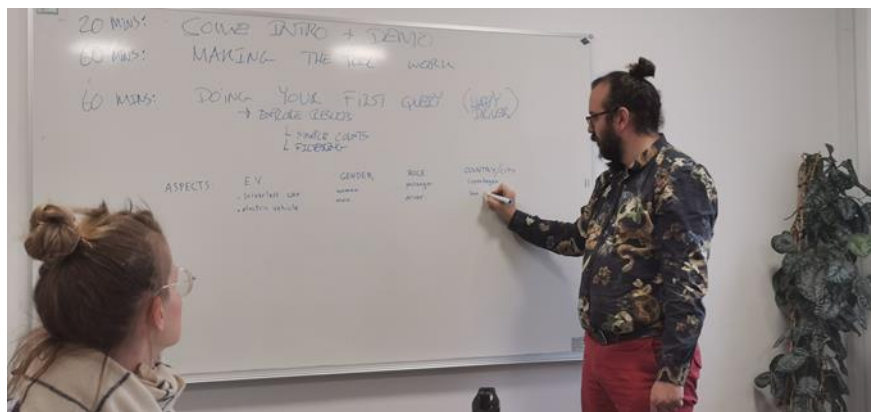
## 1. Introduction

This workshop addressed the relevance of new methodologies and big data in transport and mobility studies. Visual images and culture have been widely recognized as powerful means of how social life happens and have been staged as a “visual turn” in cultural and political analysis. Due to global media and digital means of communication, we are confronted with worldwide media events in a daily and eternal stream also when it comes to transport and mobility. Therefore, images and visual communication are important to study. Following the TInnGO grant agreement ( DOI, annex B: pp 23-24 ) the aim of the TInnGO digital/ visual analysis was to direct the focus of digital visual analysis towards smart mobility and employment, and furthermore to track both visual representations and their intersections with gender and diversity. More specifically this report, and the subsequent TInnGO digital analysis aims to explore critical aspects of power and inequality in the current visual “google world”. An image world which meets the average EU citizens when it comes to transport and mobility. To meet this purpose, we developed a new digital methodology and a digital tool to connect gender to transport and mobility in various countries in Europe. The aim was at this initial stage to introduce basic techniques and procedures and to explore the potentials of how to apply new technologies of data collection and visualization addressing gender and smart transport.

As for the findings from the various TInnGO partners they showed that the TInnGO focus on the field of gender smart mobility at one at the same time made up an appropriate field for explorative analysis of potentials and limitations: as regards the potentials, the workshop provided preliminary analysis which can be advanced and used in both transport and mobility studies. As regards limitations in current visual data collections methods they also came to the fore; mainly due to the poor visual abilities of the tagging/ translations of images into text such as personal attributes and diversity (elaborated in the TInnGO D.4.6 [https://www.tinn.go.eu/wp-content/uploads/2020/11/TInnGO\\_D4.6\\_Media-Analysis-Report.pdf](https://www.tinn.go.eu/wp-content/uploads/2020/11/TInnGO_D4.6_Media-Analysis-Report.pdf)).

## Digital workshop

The workshop was set up and hosted by University of Copenhagen/TInnGO partner in co-operation with experienced researchers from University of Aalborg in Denmark. At several preparatory meetings, feasibility and possibilities were explored and discussed by UCPH and the Aalborg team. We agreed on the provision of a shared toolbox, which could be applied to relevant key words/ search words by the TInnGO partners.



*Prior to the workshop several meetings were conducted by Ucop and Aalborg university partners in order to develop a feasible format for a one day workshop with the TINNGO partners, resulting in the conduct of a successful workshop.*



## 2. Conduct of the workshop

The workshop was conducted by Anders Kristian Munk and Mathieu Jacomy from Aalborg University, Denmark. Both are highly skilled in digital analysis and able to address an audience such as the TInnGO partners with very different scholarly backgrounds and skills.

***Anders Kristian Munk** is associate professor at Aalborg University in Copenhagen and head of the Techno-Anthropological Laboratory. He has an MSc. in European Ethnology from the University of Copenhagen and a Ph.D. in geography from the University of Oxford. Anders researches and teaches the use of machine learning and large amounts of unstructured data for cultural science studies. He is also co-author of a Danish-language textbook on digital methods.*

***Mathieu Jacomy** is a doctorate student in techno-anthropology at the Aalborg University TANTLab. He was research engineer for 10 years at the Sciences Po médialab in Paris and is a co-founder of Gephi. He develops digital instruments involving data visualization and network analysis for the social science and humanities. His current research focuses on visual network analysis, digital methods and issue mapping. He contributes to developing the web crawler Hyphe, the online network sharing platform MiniVan.*

**Focus and scope:** The focus and scope of the workshop was framed as a training workshop. In discussion with Aalborg University it was decided to focus on image tagging. During the preparatory meetings several pilot searches were conducted with combinations of various key words/ query words. The ones selected as relevant for the attendees and likely to yield a significant amount of images were ‘autonomous cars and passenger’/ and ‘employment and gender’.

Besides we wanted to present a handsome and feasible tool that could be conducted by TInnGO partners, with very different disciplinary backgrounds and knowledge of

digital methods. Since digital analysis is a tool for analyzing big amounts of data, image taggings were selected as the focus. Another reason was to enable working across geographical borders in terms of image/and language tagging, an approach that was refined and demonstrated in the subsequent targeted analysis to be found in the TInnGO D4.6 report. In the workshop each hub worked with 500 images related to each of chosen themes.

**Keywords in the digital queries:** In the preparatory phase, it was decided to transpose the smart discourse into a general query of the two themes:

1. *autonomous cars and passengers.* <sup>1</sup>
2. *transport and employment.* <sup>2</sup>

Intersectionality and the explicit search for autonomous cars and ethnicity was omitted due to the absence of relevant images in pilot queries, meaning that only white persons turned up in the trial searches. Consequently the searches were conducted along implicit intersectional guidelines, and conducted as an open ended categorical search ( implied in passenger and employment) a search which in this case yielded white gender/men and women only. <sup>3</sup>

---

<sup>1</sup> This also meant that we omitted analysis of car firms as stated in the Wp 4.4. task description, due to the lack of a substantial/quantitative cluster of digital materials from producers such as Volvo, IBM and Fiat on autonomous cars in September 2019. Tingo partners lead by Ucop will in 2021 in order to advance the method set up relevant analysis and comparisons of single firms by means of the digital tool provided.

<sup>2</sup> Employment forms a parallel key concern of the TInnGO project The tool will accordingly be explored in 2021 with a specific focus on gender and diversity in entrepreneurship in smart mobility.

<sup>3</sup> An open-ended method in intersectional research suggested by Hankivsky 2005 who argue for a diversity mainstreaming framework that transcends the current focus and forms of gender mainstreaming into a more fluid and open form of mainstreaming, where gender is kept in mind, but

## **2.1 TInnGO Digital analysis workshop.**

### **Program and introduction by the team from Aalborg University:**

“In this workshop we will explore a method for automatically tagging images from the Google Image archives, will use machine learning to predict concepts (i.e. topics and content) in images of driverless cars and do facial recognition to assess gender and diversity of the people depicted. We will learn how to repurpose Google Images to build interesting corpora of online content for comparative analysis. And we will conduct visual network analysis to understand how concepts tend to co-occur in the same images. Participants will be introduced to digital tools like Clarifai, Gephi, Table2Net and the Google Image Scraper, and we will learn to use a purpose-built script designed for the TInnGO project to help researchers leverage the relevant features of these tools in an easy to manage workflow in their internet browsers.” (Anders Munk and Matheu Jacomi.)

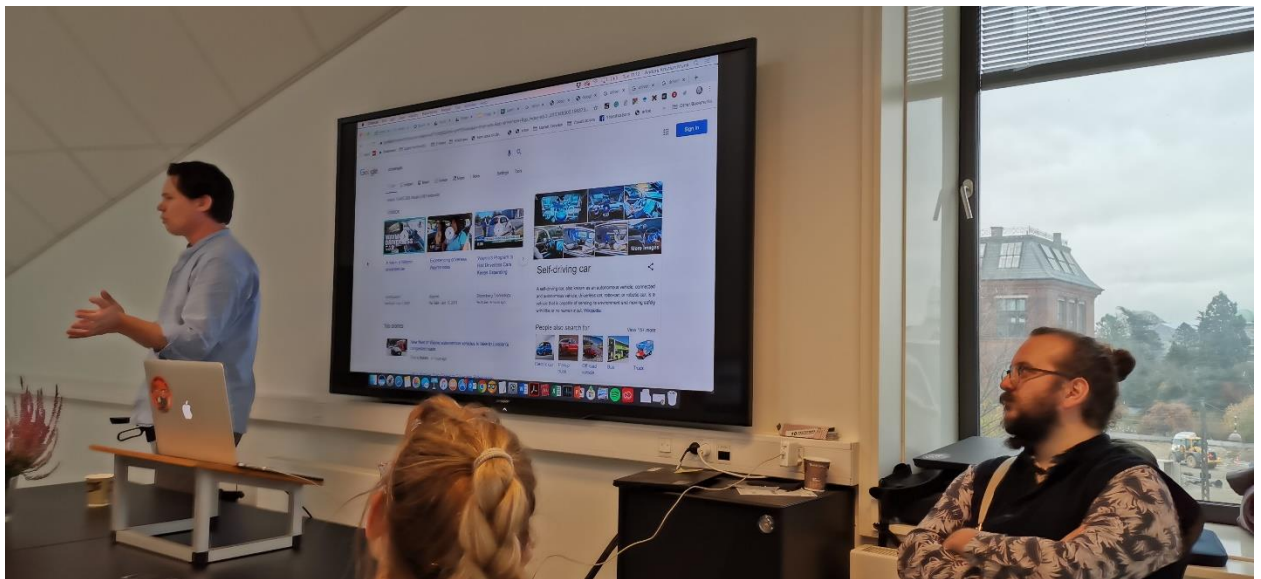
Technical details of TInnGO digital analysis – see guidelines in enclosed powerpoints by Anders Munk and Matheu Jacomi. These were supplemented by a general introduction, process guide and instructions of open access principles (enclosed as Appendix 1, 2, 3).

---

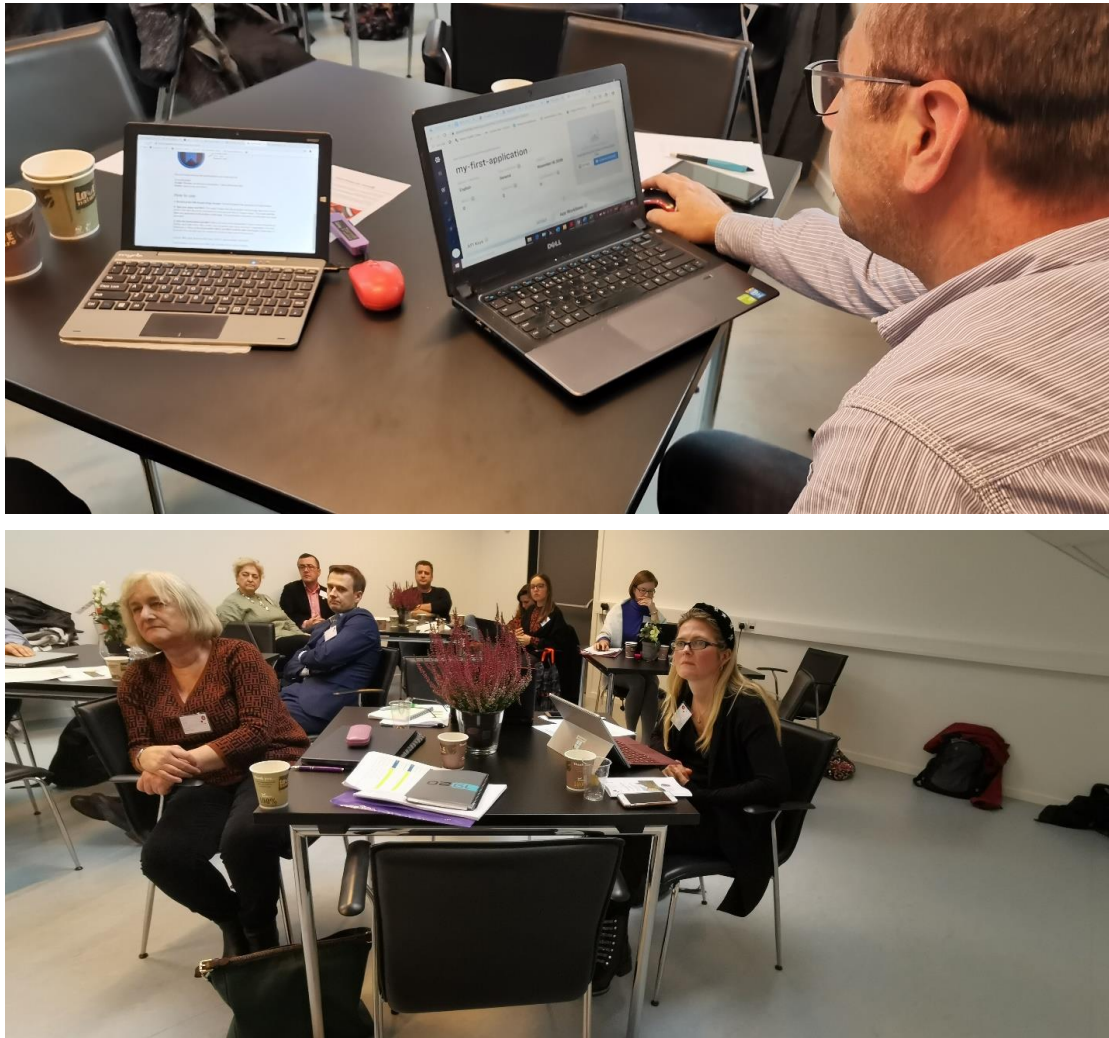
not a priori placed as the most vital category. Gender vs. Diversity Mainstreaming: A Preliminary Examination of the Role and Transformative Potential of Feminist Theory  
December 2005, Canadian Journal of Political Science 38(04):977 - 1001 DOI:  
[10.1017/S0008423905040783](https://doi.org/10.1017/S0008423905040783)

## 2.2 Program

- 11:00 – 11:20: Welcome + introduction to the methods
- 11:20 – 12:00: Get the tools up and running on your machines
- 12:00 – 12:30: Lunch
- 12:30 – 13:00: Do your first query and explore results
- 13:00 – 15:00: Build a better query design + collaborative analysis
- 15:00 – 15:30: Coffee
- 15:30 – 18:00: Visual network analysis



For the detailed content and process of the Digital workshop – please consult the powerpoint presentation inserted at the end of this report as appendix 1,2,3. These show the guided introduction and the various steps along which the digital experts offered a hands on and practical guidance for all TInnGO partners. Most exercises were conducted by members of the same hubs/ and or by ad hoc teams, and members who did not belong to a hub/or for other technical reasons of inspiration and support. Following the workshop, the partners were asked to answer a simple questionnaire, and thereby assess the usefulness and to forward their findings.



### **3. Summary of the Workshop**

In general the TInnGO partners and hubs found the digital tools and methods stimulating and useful for their future work. All partners were open to explore the possibilities of more specific searches/ queries.





### **TInnGO Digital analysis workshop – assessments:**

After the workshop all TInnGO partners and hubs were asked to provide a short report and responses of the following questions:

- a. the concrete search you made during the workshop
- b. your findings/ if you did not succeed in providing findings - explain why.
- c. your plans for incorporating digital visual analysis in your hub projects
- d. any other comments and ideas.

In the following section, extracts from the reports are inserted – according to search processes, findings and general assessments.

### **3.1 UK Hub: Gender and diversity in Smart Mobility**

- a) The search conducted during the workshop was for gender representations of smart mobility (SM). Following the discussions on the conceptualisation of what SM is and how people define it, we were interested in its visual depiction relating to gender, diversity and ethnicity.

- With less than 5% human figures in our database, ‘no persons’ was a concept appearing almost 50% of the time, depictions of SM are of futuristic, technologically integrated and fast environment with no humans in it. Words like element, graphic, vectors, symbol, particle, network, science, technology, space, speed, blur, modern, architecture, skyscraper, cityscape, futuristic, data and communication, where commonly found concepts:



- c) Following the workshop, the UK Hub will be utilising digital visual analysis to:
  - i) compare gender and diversity representations and discourses in transport and smart mobility between the US and the UK;
  - ii) understand transport and future vehicles design projects of students in the field;
  - iii) better understand how SM is socially defined and what its depiction informs us about gender, diversity and mobility.
- d) We intend to explore the tools and software provided to complement our social media analysis, as well as the possibility of video analysis.

### **3.2 Baltic Hub: Lithuania-UK comparison on future mobility**

- a) The Baltic hub investigated what *future mobility* means in Lithuania (by using .lt) and United Kingdom (by using .uk) during the workshop. The aim was to understand the perception of future mobility differences in the media in these countries.
- b) We found two different meanings of “future mobility” in Lithuanian media and in UK media. In Lithuanian media, future mobility was mostly related to modern, electric road vehicles, just a bit of autonomous cars. While in the UK media future mobility was related to much wider part of transport sector. The results included pictures with railways, digital spatial transport planning etc.
- c) The tool might be used for understanding employees in mobility and transport sector in different countries. The aim would be to see how often women are shown as employees in the transport and mobility sectors in different countries. Another option for search would be simply searching for queries “women in transport”, “women in mobility”.

### **3.3 Romanian Hub: Romania-Denmark comparison, Women in Management**

During the GDDA workshop we tried a couple of searches related to the Romanian hub specifically, such as “*women involved in management of transportation sectors, a*



*Romania and Denmark comparison*". We managed working with Clarifai and google vision/ google image scraper. Translated our design into Google queries generated a not surprising result. Set as limited by the program to 100 answers, the most relevant finding was **2 women in RO and 27 in DK**. Comparing this with the data that we already had, collected from our real research in the TIngIDLab, related to the local and county of Alba transport companies, it is not far but also not close to the truth (we have 5 women in transport management only in Alba County, so a result like " 2" for the whole Romania it's not quite accurate... ).

On the other hand, even if we have installed Java, as required, the results could not be loaded into the created network. We don't have an explanation for this, maybe it is an technical- advanced one...

### **3.4 UCPH and IR Hub: Romania-Denmark comparison, employment and technology**

To examine the visualization of the developing transport sector, we made a search on 198 relevant images in Google Image, respectively 98 images from Danish websites, and 100 images from Romanian websites. The exact search words were: "transport AND technology AND employment AND site:.dk", and "transport AND technology AND employment AND site:.ro". With this material we wished to use the Clarifai recognition service to treat the differences in gender representation in respectively Denmark and Romania.

We found the unexpected result, that human actors were not widely represented in the visualization in transport, technology and employment in neither Denmark or Romania. On the Romanian websites only 14 % of the images were registered as containing "masculine", where surprisingly 25 % of the images were registered as containing "feminine". Of the total representation of the images on the Romanian websites only 34 % were registered as either masculine or feminine, which left a great amount of images, where people were not present in the representation of transport, technology and employment. Denmark had even fewer representations of human actors represented in the visualization as only 5 % were registered as either masculine or feminine in total, consisting of only 2 % images registered as masculine, and 3 % registered as feminine.

## 4. General Assessments and Potentials

### 1.1 The Romanian Hub

The second day of the TInnGO meeting in Copenhagen, Denmark, was the most productive for the subsequent development of a hub, in general and of the Romanian hub, in particular. The workshop with the theme **Gender and diversity in digital analysis** offered a modern, practical and concrete way, unknown until then, of making effective researches in a new, fast, easy and professional way, in order to obtain data for each TInnGO hub. Considering all this, digital analysis could be a very useful tool, with the must-do condition to develop the technical skills of the involved people to obtain the most relevant results. Must be considered that it will not be able to replace efficient research, based on official data, obtained from the field.'

### 1.2 German and French Hub

The idea behind this tool in itself is very good in its approach. It offers the opportunity to find out from a flood of images and information those that are supposedly relevant to the seeker. We deliberately use the word "supposedly" because the basis for the tool is not beyond doubt. As the workshop leaders told us at our request, it is completely unknown how the tags assigned to the images come about. Therefore, one must always assume that the search contains a certain amount of errors, since the assignment of the tags can already exclude a considerable number of images that would nevertheless have been relevant for the searcher. Added to this are those that, despite the tags, do not show what they identify or the context has been misinterpreted. Therefore we would certainly use the tool, but never as the sole tool. We would always use different strategies and tools to check the results. For example, if you search for "women as employees in public transport" with Google as well, you will be shown pictures of women who actually match this search query both in terms of representation and content.

### 4.3. Mixed team (Inter + UCop)

When speaking of the approach in general terms it can give an overview quickly, and it can offer explorative insight in the visualization of a given theme or subject, e.g. we became aware of the absence of persons and faces in the images. The exploration and representation of the images are limited by the available data on the platform Google Image, and also on the reliability on the Clarifai recognizing service, as well as the restriction of categories, that the service offers. With accurate categorizations, the quantitative explorative approach can be helpful, as it is argued that e.g. explorative factor analysis is good to examine data when it is complex, and when it is difficult to conclude which factors are the most important (Kline 1994: 10).

#### **4.4 Greek Hub**

At this stage it is unknown how the digital visual analysis could be incorporated in our hub work but the aforementioned concept could be enlarged in order to reveal to local stakeholders the media or public perception of different matters in relation to transportation and gender issues.

## **5. Summary: Digital Analysis as a learning process**

### **5.1 Widening and focusing: Search topics and findings**

Most searches which were set up by the TInnGO partners focused on visual representations of smart mobility and gender, mainly women in various modalities and on employment and visualizations of women in STEM educations and vocations. These topics will be elaborated and focused in the upcoming digital analysis conducted by a sample of the TINNGO partners. A particular focus of attention will be how to maintain and elaborate the diversity component of the project including the handling of various categories such as gender, age, ability, geography. Based on the preliminary findings from the workshop the geographical components and comparisons turned out as hard to assess.

### **5.2 Technical knowledge and supervision**

As for the technical provisions most partners succeeded in installing the search tools and to follow the first vital steps in the guided process supervised by the very able team from University of Aalborg. Most participants succeeded in making various/relevant queries, and to extract image URLs while the next steps of building networks of tags in table 2 nets and in particular the visualizations and exploration network in Gephi turned out to be a challenge. Given the very compressed program and time of the digital workshop, such weak points need to be qualified/lifted in the upcoming skype supervision: Not least the translation/positions of findings into graphics and other visuals turned out to be a challenge for most parties.

The digital tool was found/ judged by the TInnGO partners to be refreshing and as a smart tool for throwing new light on relevant TInnGO themes. The TInnGO partners in their various searches also provided evidence of the tool as useful for the assessments of various human/technical intersections (the overrepresentation of technical devices and general under representation of gendered humans) comparisons across national border / various sectors etc. As such the digital tool and methods were regarded as having potential given the scope of the project. At the same time several partners found, that the findings have to be analyzed carefully and made subject to

further validation – e.g. through sample control of visuals and their contents. It has to be stressed that the digital analysis cannot stand alone, but can be seen as a useful and stimulating supplement to other research methods, such as the use of conventional statistics and interviews and mappings related to various transport modalities and practices.

In conclusion, the digital workshop ought to be seen as the starting point in a learning process of how digital methods can be applied in new and path-breaking ways in relation to a more inclusive and dynamic transport and mobility research. Follow up activities will be conducted by a sample of TinnGO partners primo 2020 – and TinnGO hubs may join this process. At the same time local teams – at both hub and partner levels are urged to explore the tools provided further at the local levels – where additional technical assistance might be available.

**Spring 2020: follow-up activities of TinnGO, Digital (media) analysis.**

TinnGO partners involved in the upcoming digital analysis deliverable 4.6 are SOCTR/ UK, CU/UK, VTI/SWED, VTM/Port/, INTECO/ Rou, where UCPH is lead. The cut-off date for the deliverable 4.6 Media Analysis was 1st May 2020.

## Appendix

### 1. TInnGO digital analysis

TInnGO digital – see guidelines in enclosed powerpoint presentation and process guide. All files for the workshop can be found here:

<https://docs.google.com/presentation/d/1Q4LPwrxUt6b3aldqJzB9YTK1cSn7o1ySr dKkXuJvpc/edit?usp=sharing>

### 2. Additional guidelines

Additional guidelines For the tool itself, the steps are on the web page. You'll find steps for the setup on top, and steps for the use on the bottom ("How to use").

<https://jacomy.github.io/google-image-dmi-clarifai/>

For the network of tags, I have shown in the end, the steps are:

1. Upload the CSV to Table2Net
2. Follow the instructions of Table2Net to build a network where nodes are the tags from column "general concepts" and the links are the row number (i.e. when two tags qualify the same image)
3. Download the GEXF file from Table2Net (it's a network)
4. Open the GEXF with Gephi and visualize it to explore its topology

For using Table2Net (to make a network of tags), the steps are also featured inside the tool (it will appear on the right side of the screen once you upload a CSV)

<https://medialab.github.io/table2net/>

For a step-by-step use of Gephi, you can follow this tutorial:

<https://medium.com/@EthnographicMachines/visual-network-analysis-with-gephi-d6241127a336>

### **3. Copyright and Open Access**

About the license. For the sake of clarity, although I am sure you already get it: Clarifai itself is just a service we use, it is not inside the tool itself. The data we get from Clarifai (the tags) can be used freely.

The tool itself is licensed in GPL v3, the most common open source license. It guarantees that anyone can freely download it, copy it, modify it, reuse it, and even sell it. The only condition is that if it gets used as a part of another software, that one must also be under that license (legal teams know how it works, it's a classic). The source code of the software is published on the GitHub platform:  
<https://github.com/jacomyma/google-image-dmi-clarifai>

That's what you want to download if you want to save a copy.

The license is part of that repository, technically you can find it by clicking on LICENSE or by using that link:

<https://github.com/jacomyma/google-image-dmi-clarifai/blob/master/LICENSE>

The website, which is the part visible to the users (contrary to the source code) is hosted on GitHub. You can host it on your own if you want, no problem. It is a static web page, nothing difficult here:

<https://jacomyma.github.io/google-image-dmi-clarifai/>

### **4. Gender and diversity in Digital Analysis**

*On the next pages you will find the slides, that were presented during the workshop.*

---

## Gender and diversity in digital analysis

*TINNGO WORKSHOP*

Anders Kristian Munk & Mathieu Jacomy  
The Techno-Anthropology Lab (@tantlab)  
Aalborg University Copenhagen



## Program

11:00 – 11:20: Welcome + introduction to the methods

11:20 – 12:00: Get the tools up and running on your machines

12:00 – 12:30: Lunch

12:30–13:00: Do your first query and explore results

13:00–15:00: Build a better query design + collaborative analysis

15:00–15:30: Coffee

15:30–18:00: Visual network analysis

---

---

# Introduction to methods

---



---

## 3 methods

1. Repurpose Google for research
  2. Use machine learning to automatically tag large volumes of images
  3. Use visual network analysis to identify clusters of co-occurring tags
- 

---

## Studying (stuff with) Google

**Repurposing:** What can we learn from studying the way platforms and search engines organise the world?

**Media effects:** How do we show and handle media bias?

**Search is research (Rogers 2009):** How can we construct queries that make interesting differences visible? E.g. differences between geography, time, search terms, content type, websites.

---

---

# Experiment!

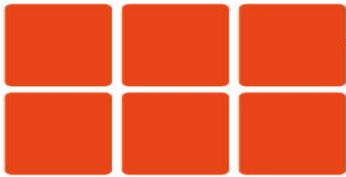
---



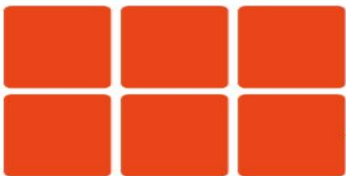
Car  
Face  
Male  
Keys

Image tagging

Training set of manually tagged images

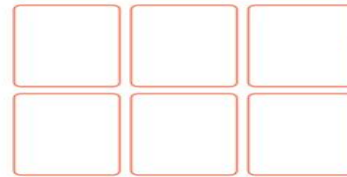


Test set of manually tagged images

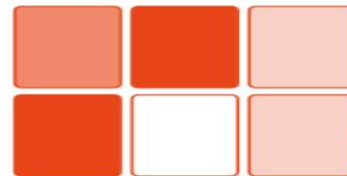


Train model to recognize visual patterns corresponding to tags  
**>>> validate**

Set of untagged images



Set of tagged images with different levels of confidence



Use model to tag images

clarifai

PRODUCTS - ENTERPRISE - DEVELOPERS - COMPANY - DEMO - PRICING - LOG IN



**Color**  
Identify the dominant colors present in your media in hex or RGB form



**Demographics**  
Predict the age, gender, and cultural appearance of detected faces



**Face Detection**  
Detect the presence and location of human faces with a bounding box



**Face Embedding**  
Computes numerical embedding vectors using our 'Face Detection' model



**Food**  
Recognize food items and dishes, down to the ingredient level



**General**  
Our most comprehensive model with concepts, objects, scenes, and more

Clarifai models (<https://www.clarifai.com/models>)

---

## Challenges

**Algorithmic bias:** The machine is never better than the training set.

**Proprietary software:** Easy of use = black boxed algorithms and training sets

**Unexplainable AI:** Even with full transparency and supervised machine learning (known training sets) it can be difficult to explain exactly why non rule-based algorithms predict the tags they do

---

---

## Benefits

**Volume:** We can quickly evaluate the content and spot patterns in large batches of images.

**Ease of use:** Clarifai or Google Vision, although proprietary and somewhat opaque, offers nice user interfaces and well-documented API's.

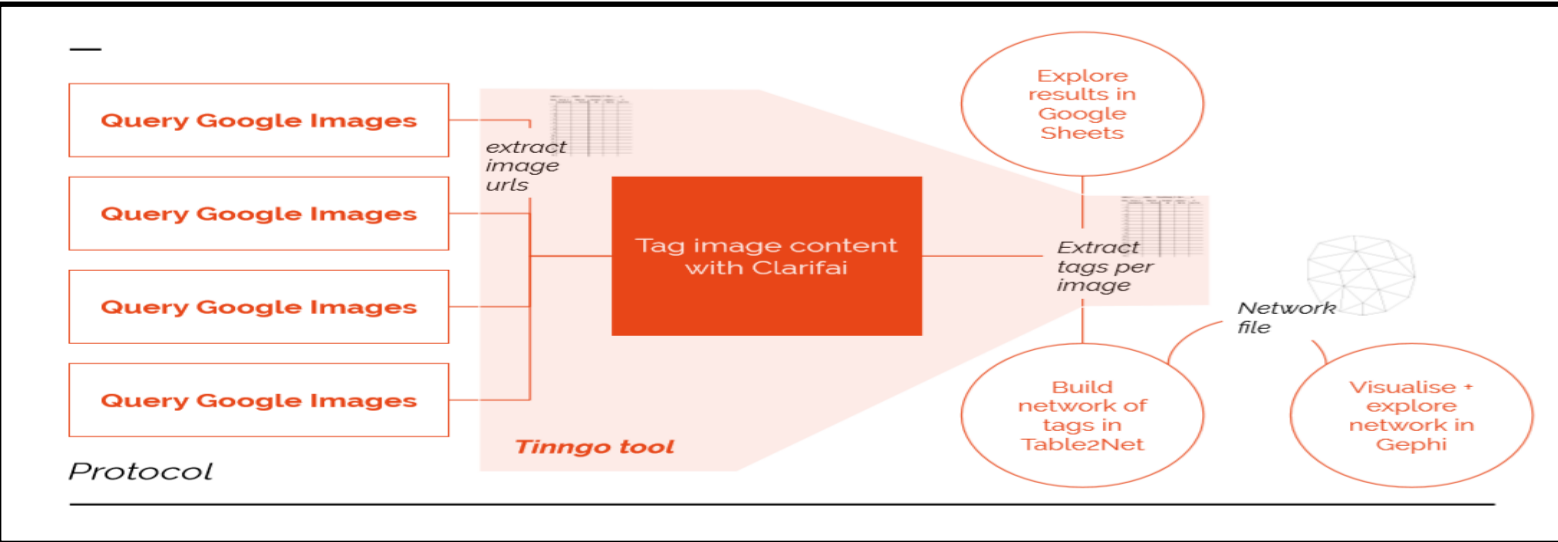
**Interface with qualitative methods:** We can easily move from patterns and statistics to a qualitative content analysis of individual images

---

---

**Again: Experiment,  
experiment, experiment...**

---



## Query design

## Objective

- Systematically conduct a series of searches on Google Images that will be interesting to compare
- Leverage Google search operators and syntax
- Distribute the searches in your group (to exploit multiple free trial accounts) and collect the outputs from Clarifai

### ASPECT #1



#### GOOGLE QUERIES

### ASPECT #2



AND

### ASPECT #3



### ASPECT #1: *mode of transport*

*driverless car*  
*electric bike*

#### GOOGLE QUERIES

"driverless car"  
"driverless car"  
"driverless car"  
"driverless car"  
"driverless car"  
"driverless car"

"electric bike"

...

### ASPECT #2: *situation*

*commute*  
*commercial*  
*leisure*

AND commute  
AND commercial  
AND leisure  
AND commute  
AND commercial  
AND leisure

AND commute

...

### ASPECT #3: *role*

*driver*  
*passenger*

AND driver  
AND driver  
AND driver  
AND passenger  
AND passenger  
AND passenger

AND driver

...



## ASPECT #1: mode of transport

driverless car  
electric bike

### GOOGLE QUERIES

"driverless car"  
"driverless car"  
"driverless car"  
"driverless car"  
"driverless car"  
"driverless car"  
"electric bike"  
...

## ASPECT #2: situation

commute  
commercial  
leisure  
syn: travel

AND commute  
AND commercial  
AND (**leisure OR travel**)  
AND commute  
AND commercial  
AND (**leisure OR travel**)  
AND commute  
...

## ASPECT #3: role

driver  
passenger

AND driver  
AND driver  
AND driver  
AND passenger  
AND passenger  
AND passenger  
AND driver  
...

## ASPECT #1: mode of transport

driverless car  
electric bike

### GOOGLE QUERIES

"driverless car"  
"driverless car"  
"driverless car"  
"driverless car"  
"driverless car"  
"driverless car"  
"electric bike"  
...

## ASPECT #2: situation

commute  
commercial  
**not: advertising**  
leisure  
syn: travel

AND commute  
AND (**commercial -advertising**)  
AND (leisure OR travel)  
AND commute  
AND (**commercial -advertising**)  
AND (leisure OR travel)  
AND commute  
...

## ASPECT #3: role

driver  
passenger

AND driver  
AND driver  
AND driver  
AND passenger  
AND driver  
AND passenger  
AND driver  
...

Aspect #1

Aspect #2

Aspect #3



Filter #1  
Country



Filter #2



Filter #3

Denmark  
France  
Germany  
U.K.  
Netherlands

### Queries:

"driverless car" AND commute AND driver **AND site:.dk**  
"driverless car" AND commute AND driver **AND site:.fr**  
"driverless car" AND commute AND driver **AND site:.de**  
"driverless car" AND commute AND driver **AND site:.co.uk**  
"driverless car" AND commute AND driver **AND site:.nl**



## Queries:

"driverless car" AND commute AND driver AND site:.dk  
 "driverless car" AND commute AND driver AND site:.fr  
 "driverless car" AND commute AND driver AND site:.de  
 "driverless car" AND commute AND driver AND site:.co.uk  
 "driverless car" AND commute AND driver AND site:.nl

+ Time setting in the Google Image Scraper



## Queries:

"driverless car" AND commute AND driver  
 "driverless car" AND commute AND driver  
 "driverless car" AND commute AND driver  
 "driverless car" AND commute AND driver  
 "driverless car" AND commute AND driver

+ URL search box in in the Google Image Scraper

## Your task

1. Come up with a concept for an interesting query design. Essentially: What are the differences we want to be able to compare?
2. Translate that design in to actual Google queries
3. Discuss your proposed list of queries with Anders or Mathieu and revise
4. Distribute the queries in your group and use the Google Image Scraper + the TInngo Tool to download data files with tags for each query

**TinnGO** is funded by the European Horizon 2020 program, and its objective is to promote gender equality and diversity in the transport sector in the European Union. The project, led by **Coventry University**, began in December 2018 and will run for three years.



POLITECNICO  
DI TORINO



CITTÀ DI TORINO



## Principal investigator

Andree Woodcock

adx974@coventry.ac.uk



TinnGO



TinnGo



@TinnGoEU



[www.tinnGo.eu](http://www.tinnGo.eu)



This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 824349

**Disclaimer:** This document reflects only the author's view. Responsibility for the information and views expressed therein lies entirely with the authors. The Innovation and Networks Executive Agency (INEA) and the European Commission are not responsible for any use that may be made of the information it contains.