University of Zagreb
FACULTY OF HUMANITIES AND SOCIAL SCIENCES

Beatrice Züll

SOCIAL MEDIA ACTIVITY IN THE CONTEXT OF FREE-TO-AIR BROADCAST TELEVISION

DOCTORAL DISSERTATION

Zagreb, 2017

University of Zagreb
FACULTY OF HUMANITIES AND SOCIAL SCIENCES

Beatrice Züll

SOCIAL MEDIA ACTIVITY IN THE CONTEXT OF FREE-TO-AIR BROADCAST TELEVISION

DOCTORAL DISSERTATION

Zagreb, 2017

Sveučilište u Zagrebu
FILOSOFSKI FAKULTET

Beatrice Züll

ULOGA DRUŠTVENIH MEDIJA U MJERENJU TV GLEDANOSTI

DOKTORSKI RAD

Zagreb, 2017.

University of Zagreb
FACULTY OF HUMANITIES AND SOCIAL SCIENCES

Beatrice Züll

SOCIAL MEDIA ACTIVITY IN THE CONTEXT OF FREE-TO-AIR BROADCAST TELEVISION

DOCTORAL DISSERTATION

Mentor:
Mihaela Banek Zorica, Ph.D.

Zagreb, 2017

Sveučilište u Zagrebu
FILOSOFSKI FAKULTET

Beatrice Züll

ULOGA DRUŠTVENIH MEDIJA U MJERENJU TV GLEDANOSTI

DOKTORSKI RAD

Mentor:
izv. prof. dr. sc. Mihaela Banek Zorica

Zagreb, 2017.

Acknowledgements

This doctoral thesis was supported by the members and colleagues of the Faculty of Information and Communication Sciences of the University of Zagreb who provided insight and expertise and greatly assisted the research. I would like to express my sincere gratitude to my supervisor Mihaela Banek Zorica, Ph.D., for her guidance, encouragement and expertise. Thanks to my advisor Stephan Malović, Ph.D., for his encouragement and great support.

Alida Zorz Miketek from Nielsen supported this research work as an expert in television audience research. Guillaume Antoine Rabi and his team at RTL Televizija Hrvatska endured the challenges of the empirical part of this project by giving advice and insights into empirical data. Poslovna Inteligencija, Dražen Oreščanin and his team supported this project with staff and server capacity. Fremantle Production and Ana Habajec provided insights into the strategic approach of television production companies to integrating social media in their daily work and building up social media audiences.

Thank you to all those in the broadcasting industry who kindly gave their time and advice, I am extremely grateful. Finally, I thank my family and friends who provided support through the process of researching and writing.

Thanks to Croatia.

Beatrice Züll

Mentor

Mihaela Banek Zorica has graduated informatology and Czech language and literature at the Faculty of Humanities and Social Sciences. In 2007 she has defended her Ph.D. in the field of Information sciences with the title Systems for management of educational materials in electronic environment. Since 2012 she has been Chair for media and communication. Her teaching is on knowledge organization and management, e-learning, school libraries and information literacy on undergraduate, graduate and postgraduate level. She has participated on number of national and international conferences and seminars as an author, reviewer or member of the organizational or program board. Her research was done on several international and national projects: Organization, management and knowledge sharing in electronic educational environment, Organization of information and knowledge in electronic environment, TEMPUS Joint European Project - Aspects of Organization and Information Systems: Curriculum Development, TEMPUS TEALS (Teaching and Language skills). In 2011, she was coordinator of ERASMUS Intensive program: Information and communication technology in supporting the educational process.

In 2010, she has received two awards: University of Zagreb Award for best utilization of social software in e-learning, and the IASL School Librarianship Award.

Mentor

Mihaela Banek Zorica diplomirala je informatologiju i češki jezik i književnost na Filozofskom fakultetu Sveučilišta u Zagrebu nakon čega upisuje Poslijediplomski studij informacijskih znanosti. Godine 2007. doktorira u području informacijskih znanosti obranivši rad pod naslovom Sustavi za upravljanje obrazovnim materijalom u elektroničkom okruženju. Od 2012. godine predstojnica je Katedre za medije i komunikologiju. Predaje na preddiplomskom, diplomskom i doktorskom studiju kolegije vezane uz organizaciju znanja i upravljanje znanjem, e-učenje, školske knjižnice i informacijsku pismenost. Sudjelovala je na nizu međunarodnih i nacionalnih konferencija i skupova u vidu autora znanstvenih radova, recenzenta i člana organizacijskog ili programskog odbora. U koautorstvu je objavila dvije knjige, poglavlja u knjigama te više od 30 znanstvenih i stručnih radova.

Sudjelovala je na nekoliko međunarodnih i nacionalnih projekata: Organizacija, upravljanje i razmjena znanja u elektroničkom obrazovnom okruženju, Organizacija informacija i znanja u elektroničkom obrazovnom okruženju, TEMPUS Joint European Project - Aspects of Organization and Information Systems: Curriculum Development, TEMPUS TEALS (Teaching and Language skills). Godine 2011. koordinator je ERASMUS Intenzivnog programa Information and communication technology in supporting the educational process.

Godine 2010. dobitnica je dviju nagrada: Nagrade za uporabu društvenog softvera u e-kolegiju Sveučilišta u Zagrebu, Povjerenstva za e-učenje za najbolji e-kolegij za akademsku godinu 2009/2010., te godišnje nagrade The IASL School Librarianship Award.

SUMMARY

For many decades, traditional broadcast has been the main entertainment focal point in households. Like all media and entertainment industries, television has been altered by the internet and new technologies. The internet has made new forms of participatory communication possible and has increased the amount of interpersonal communication for individuals – audiences and users – providing opportunities to share, create and collaborate together. It offers manifold opportunities to communicate in all directions, as well as the opportunity to transmit and receive simultaneously all kinds of content and formats such as music, films, pictures and texts and enables the user to interact with links. The development of social media is more than a technical innovation: it sustains and influences all forms of social organisations. Besides (high speed) internet itself, wireless connectivity has created a comfortable environment for the usage of different devices. Smartphones, tablets and/or laptops are conquering households and invite (connected) usage while people watch TV; audiences divide their attention between a second and first screen, becoming a user and audience at the same time. It enables participation and social interaction within social media while watching TV. "Actions of the participatory audience appear in the value chain in several phases: when the audience is creating content, when they are editing or reediting the available content and when they are disseminating the content to other audience members" (Noguera Vivo et al., 2014, p. 181). This new participation of TV audiences in social media leads to an integration of TV consumption within the social media context. The "people formerly known as the audience are those who *were* on the receiving end of a media system that ran one way, in a broadcasting pattern, with high entry fees and a few firms competing to speak very loudly while the rest of the population listened in isolation from one another […]" (Rosen, 2006), the audience transformed into an active audience participating in the creation of (social) media content.

The second screens enable virtual communication with friends about programs while watching and sharing what is liked and disliked, and television viewing coupled with audience interaction has gained popularity (Doughty, Rowland and Lawson, 2011). The audience can share, discuss, comment and vote about certain programs. Broadcasters and other suppliers offer applications accompanying TV consumption and solicit

simultaneous usage. Audiences engage with the program and socialise with friends and communities around their favourite content. Television audience researchers discovered the internet as a source of audience data, and search for approaches to analyse online engagement of audiences. The main question of this work is to investigate if new data can be found and used in a systematic manner in addition to traditional television audience research methods. It was discovered that the relationship between television broadcasters and its social audience is the key to this approach. Traditional media such as TV broadcasters are still huge content providers and play a major role in the social media world, where content is shared and creates buzz and in addition users generate content themselves. Broadcasters are challenged to keep the relationship with and the attention of the viewer by building social interaction around the program. This is the prerequisite for the researcher to approach social media analysis in the context of television.

Key Words:

Social media, Twitter**,** audience research, opinion mining, people-meter-measurement, television ratings, social media analytics, social television, social TV metrics, television media.

SAŽETAK

Tradicionalno emitiranje je već desetljećima žarište zabave u kućanstvima. Kao i svi ostali mediji te industrije zabave, televizija je promijenjena zahvaljujući internetu i novim tehnologijama. Internet je omogućio nove forme komunikacije između sudionika te je povećao broj međuljutske komunikacije za pojedince – gledatelje i korisnike – tako što je omogućio prilike za zajedničkim dijeljenjem, stvaranjem i surađivanjem. Pruža mnoge prilike za komunikaciju u svim smjerovima, jednako kao i priliku za simultano slanje i primanje raznih vrsta sadržaja te formata kao što su muzika, filmovi, slike i tekstovi. Također omogućuje korisniku interakciju s web linkovima. Razvoj društvenih medija je više od tehnološke inovacije, ono podržava i utječe na sve oblike društvenih organizacija. Pored toga, sam internet (velike brzine) je uz bežično spajanje stvorio ugodnu okolinu za korištenje raznih uređaja. Pametni telefoni, tableti, i/ili laptopovi osvajaju kućanstva te pozivaju korisnika na online spajanje i korištenje interneta za vrijeme gledanja televizije pa tako gledatelji dijele svoju pozornost između dva ekrana, postajući na taj način istovremeni korisnici i gledatelji. Ovo omogućuje sudjelovanje te društvenu interakciju unutar društvenih medija tijekom gledanja televizije. "Djela uključenih gledatelja se pojavljuju u lancu vrijednosti u nekoliko faza: kada gledatelji stvaraju sadržaj, kada uređuju ili preuređuju dostupan sadržaj, te kada šire sadržaj drugim gledateljima. " (Noguera Vivo et al., 2014, p. 181). Ovo novo sudjelovanje gledatelja na društvenim medijima vodi k integraciji gledanja televizije unutar konteksta društvenih medija. "Ljudi koji su prethodno prepoznati kao gledatelji su bili na primajućem kraju medijskog sustava koji se kretao u jednom smjeru, po strukturi emitiranja, uz visoke članarine te nekoliko tvrtki koje se natječu u tome da govore što glasnije dok ostatak populacije sluša u međusobnoj izolaciji […]" (Rosen, 2006), gledatelji su se pretvorili u aktivne gledatelje koji sudjeluju u stvaranju sadržaja (društvenih) medija.

"Dodatni zasloni omogućavaju virtualnu komunikaciju s prijateljima o TV programima tijekom gledanja i dijeljenja sadržaja koji im se sviđa, odnosno ne sviđa, a i samo gledanje televizije s istovremenom interakcijom gledatelja postaje sve popularnije." (Doughty, Rowland and Lawson, 2011). Gledatelji mogu dijeliti sadržaj, raspravljati, komentirati te glasati za određene TV programe. Televizijske kuće i ostali dobavljači nude aplikacije za praćenje korištenja usluge televizije te potiču njezino simultano korištenje. Gledatelji se

uključuju u TV programe te raspravljaju s prijateljima i raznim zajednicama o njihovim najdražim sadržajima. Istražitelji koji prate gledanost televizije su prepoznali internet kao izvor podataka o gledateljima te istražuju pristupe za analizu online angažiranosti gledatelja. Potrebno je istražiti mogu li se pronaći novi podaci koji se mogu iskoristiti na sustavan način uz tradicionalne metode istraživanja gledanosti televizije. Otkriveno je da je ključ ovom pristupu sam odnos između televizijskih kuća i njihove publike, odnosno gledatelja. Tradicionalni mediji kao što su televizijske kuće se i dalje smatraju značajnim pružateljima sadržaja te igraju važnu ulogu u svijetu društvenih medija, gdje se dijele sadržaji koji stvaraju vijesti, i sadržaj pružaju sami korisnici. Izazov televizijskih kuća je da održavaju odnos s gledateljima te da imaju njihovu pozornost tako što će izgraditi društvenu interakciju oko TV programa. Ovo je preduvjet istražiteljima kako bi pristupili analizi društvenih medija u kontekstu televizije.

Ključne riječi:

Društveni mediji, Twitter, istraživanje gledanosti, istraživanje mišljenja, mjerenje pomoću peoplemetra, televizijska gledanost, analitika društvenih medija, društvena televizija, mjerenje društvene televizije, medij televizije.

Table of Content

1 Introduction 1

2 Technology changes television media 8

2.1 Digitalisation of television 9

2.2 The internet and social web 11

2.2.1 The internet as the source of big data created by television audiences 14

2.2.2 With digitalisation towards fragmentation of television 16

2.2.3 Social media and television 21

2.3 Croatian television in transformation 28

2.3.1 Characteristics of the Croatian population and general media indicators 29

2.3.2 Television landscape in Croatia 31

2.3.3 Advertising as the main financial source for commercial television 34

2.4 Audience research in transformation 35

3 Challenges in television audience research 42

3.1 Television audience analysis in context of history 44

3.2 Description of television ratings methodology 47

3.2.1 Description of people-meter methodology 47

3.2.2 Sociodemographic data, television ratings and market shares 51

3.3 Limits of traditional methodology 58

4 Researching social media activities 61

4.1 Television audience and microblogging 65

4.2 Twitter: structure and communication 67

4.3 Approaching Twitter data for audience research 74

4.3.1 Opinion mining for audience research 76

4.3.2 Twitter web archives (collection and corpus) 79

5 Empirical study 85

5.1 Discovering social television response using opinion mining tools 86

5.1.1 Discovering sentiment with Sentiment140 89

5.1.2 Twitter search with feelings: Tweetfeel 90

5.1.3 Searching the sentiment with Tweettone 92

5.1.4 Getting real time insights from Topsy 93

5.1.5 Social media search with social mention 95

5.2 Twitter web archives: Data collection corpus and sampling 96

6 Evaluation and Modell 107
6.1 Tracking popularity with social media analytic tools 113
6.2 Opinion mining working with data collections 118
6.3 Development of social media activities of television 119

7 Conclusion 127

8 Bibliography 134

Tables 173

Figures 174

APPENDIX 1 175

APPENDIX 2 176

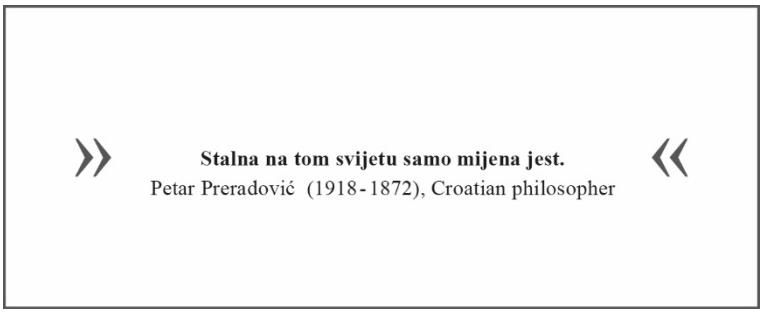
APPENDIX 3 179

APPENDIX 4 184

Curriculum Vitae 185

Životopis 187

1 INTRODUCTION



Through the processes of digitalisation and convergence, the markets of television have been in a continuous change. The number of television channels has multiplied and at the same time people consume television programs on any kind of device at any time and any place. The wireless interconnectivity enables television viewers to use second screen devices, such as tablets, laptops and mobile phones while they watch television. There is no close attention any more, but divided attention among several screens.

When a computer network connects people, just as a computer network is a set of machines, a social network is a set of people connected by a set of social relations, such as friendship, or information exchange (Garton, Haythornthwaite, Wellmann, 1997). While television used to be a social community media, where families or groups watched together, an increasing individualisation of television consumption can now be observed (Bjur, 2009) and people use social networks to communicate and express their opinion towards a television program or television station.

Social networks such as Twitter are considered as the ideal partner for big live television events as they enable the viewers to comment spontaneously and briefly and therefore reflect opinions and thoughts regarding a television program. Television as entertainment and information media has always delivered topics to talk about. The internet enables masses to talk online on social media and a high number of these comments is related to television. According to Talbot (2011), of the approximately 300 million public comments made online every day, 55 million are posted on Twitter and an average of 10 million posts a day are related to television.

Statements about numbers and shares of television related comments need to be evaluated carefully as, given the big volume of data, it is obvious that these kinds of numbers are difficult to verify. Absolute numbers in Croatia or other countries are therefore difficult to find in literature and it could be considered a legitimate question to ask how many television related comments in Croatia are generated by the audience every day.

Worldwide, millions of television viewers share their television experiences on Twitter, resulting in the creation of posts or other social media content: a library of social media data is built up which can be seen as cultural heritage of the modern society (Library of Congress, 2013) and a tool for researchers.

The analysis of social media activity as a new approach to researching television audiences has become more and more interesting for the television industry, similar to other areas of the economy where the user becomes transparent through its activities on the internet.

It is a well- known phenomenon that the increasing mobile usage of digital forms of communication is present in all areas of life and leaves digital traces everywhere.

The television viewer seems to live in a glass house where there are no secrets in respect to taste, opinions and feelings once these are shared online. The question of the measurability of success of television shows and television stations and the search for the possibilities to analyse audiences is as old as television itself. As early as the nineteen-sixties the television ratings measurement system was developed by Arthur C. Nielsen to research audiences and for decades the people-meter systems methodology has been submitting valid answers about the viewing behaviour of audiences.

The aim of television audience research has always been to better understand television audiences based on common objective methodologies and access to common data. Television audience research plays a significant role in the television industry. The provided television ratings data is a common currency for television executives, producers of television programs and the advertising industry. The traditional means of researching television audiences is measuring television ratings, which has been the primary source of information about audience behaviour since the nineteen-nineties. The television and advertising industry demands detailed data about its audiences and uses ratings to analyse the popularity of a television program and advertising. The measurement system provides information such as viewing patterns and the demographic composition of television audiences.

Worldwide penetration of internet goes hand in hand with the usage of social media platforms. The user of social media intentionally or unintentionally leaves data on the internet, depending on the social media platform provider and can be used by industry and sciences for research purposes and analysis, in particular for the research of television audiences.

Social media have created opportunities to study social data in new ways and massive amounts of data leads to a fundamentally new digital approach in Human Sciences (Manovich, 2012). Using social media for academic research is a new and growing phenomenon, as social media can be used as a key source for studying how people communicate and interact. Together with computational tools, big data can be processed and analysed and human sciences can take advantage of the new data sources (Mayer-Schönberger and Cukier, 2013). The question we now face is not how to obtain audience data but how to use data generated by the usage of social media which, intentional or not, delivers more information about audiences than the television market has ever had before.

The analysis of social media data cannot be a statistically representative approach to learning about television households in general, as the users of social media platforms represent - depending on size and constitution of the platform – only a small part of national television audiences. The more comments are created by television viewers, the more active they are on social media and the more data can be found and interpreted. There is a critical mass of data which, when it is reached, can increase the significance of data and results.

A different distribution and popularity of social networks can be observed in different countries. Facebook and Twitter belong to the most successful social media networks of the world. Croats prefer Facebook with 1.2 million users (allin1Social, 2016) but Twitter usage, although still in its infancy (Government of the Republic of Croatia, 2016), is growing with 51,986 users (Hrastovčak, 2013).

Although the usage of Twitter is small, one has to consider that Twitter data is available for research. Facebook does not give open access to their data and one of the reasons for its success is the sense of protection when users share their content with their friends. Discussing the implications and effectiveness of the protection strategy is beyond the scope of this work. However, the data situation of Facebook and its complex privacy settings significantly affect the extent to which researchers may access the data (Giglietto, Rossi and Bennato, 2012). In contrast to Facebook, Twitter is not necessarily limited to a specific group of friends, rather messages are posted to the public and information can be searched for. Twitter

has proven to be a suitable social network for research as Twitter gives access to Twitter data via the open Twitter API, which allows the researcher to track the Twitter data stream. Twitter has therefore become a widely used and researched communication channel, from politics and crisis communication to journalism (Bruns & Burgess, 2011; Bruns and Stieglitz, 2012; Boyd, Golder and Lotan, 2010) and through its use as a backchannel of television (e.g., Bruns and Stieglitz, 2012; Bredl, et al. 2014; Deller, 2011).

The more data the better, but in general which platform would be used as a source for data analysis should not play a role as long as the availability of data is guaranteed.

The methods of social media analysis and the approach to the data should, in an ideal world, be independent of the data source and should be applicable to any social media platform. Different approaches to analysing television viewers' social media activity data can be found. People share different information in the context of television consumption such as photos, links, videos and comments. They use different social media platforms such as YouTube, Facebook, Twitter or any others to exchange content in the context of television.

This research work will focus on the analysis of comments which were posted to express a thought or attitude towards television programs on social networks. When people share and discuss their opinion towards certain television programs or express their feelings one can assume that they have seen the television program. It does not mean that they like it, as there seems to be a tendency to express dislike rather than like. The search for opinions was considered to be something new compared to television ratings measurement as television ratings give no information regarding whether or not somebody liked or disliked a television program.

It will be analysed whether social media data can be used for television audience research and if ways can be found to provide reliable data and information about television audiences as an addition to existing television ratings. Social media data generated in the context of television consumption will also be researched, to determine if it can be a new valid source of information. Three hypotheses are deducted from this approach:

The first hypothesis of this research is that social media plays an increasingly important role for the television audience, giving people the opportunity to share their viewing experiences in various forms and on various platforms such as Twitter. The second hypothesis is that a relationship can be found between television consumption and social media activity.

The third hypothesis is that social media activity can impact on television ratings, and therefore influences the success of a television program. Many questions arose during this work, and methodological questions regarding how to approach the social media data dominated the research:

Is it possible to learn about television audiences by social media analytics and add information to traditional television ratings in a systematic way? A possible way to find and retrieve data will be researched and a methodological framework will be developed to approach data which allows the conclusions and generalisation from the analysis of small data and in addition the application of the framework model in other countries.

The methodological framework is divided in two sections. Data will be approached by the use of existing sentiment analysis tools (4.3.1) and by working with data corpus (4.3.2) to apply sampling as a methodical approach. It will be expected to find data which could help in learning more about television audiences and in particular the data besides the text itself which can be retrieved could be of interest to the researcher. Finally, the question that needs to be answered is what the results of this analysis could mean to the Croatian television audience research market.

The approach to work with data corpus implies to find appropriate data collections and different possibilities were researched.

In 2013 the Library of Congress in the US and Twitter published the transfer of data and ongoing archiving of the Twitter collection to the Library, taking a big step towards preserving the US digital heritage of Tweets. They announced that this Twitter *library* was going to be made accessible to researchers and policymakers in a comprehensive manner. The usage of the term *library* in the context of *Twitter* may be a further indication of the interpretation of social media data as a library for research in Human Sciences. The organisation of this data and the question regarding who can access it seems to still be open. It would be the most sophisticated approach for researchers to get open access to local Twitter data similar to the model in the US which means that national Twitter libraries archive digital heritage of Tweets.

Another approach of Twitter to providing data to the public for research purposes are data grant programs (appendix 1) which are given from to time to time to research teams. One example was the data grant program in 2014, where public research institutions were invited to

apply for data grants and get therefore get the chance to access the data. Part of this research was the application (appendix 2) to get the data grant and to have access to a Croatian Twitter data collection. As the application was declined (5.2), other possibilities had to be found to get access to a social media data collection to construct a similar scenario.

The aim of all approaches is the same: how to get access to data in a comprehensive and transparent manner so that it can be used in a transparent way for research.

Within the framework of this research, data was used which was retrieved through Twitter API. There was an awareness regarding the restrictions, and that the picture of data could not be complete. However, the focus was on the evaluation of existing social media analytic tools and methods as instruments to analyse Croatian Twitter data and also to find new ways to approach data as a web archive (4.3.2). The purpose of the analysis of Croatian Twitter data is to interpret the outcome and to gain possible insights which could be helpful to learning more about the Croatian television audience. The main instrument of the television industry for analysing the behaviour and habits of television audiences has for decades been the analysis of television ratings generated with television audience measurement based on people-meter methodology. Television ratings data are used by all participants of the television market as a standard, because every market needs valid and reliable data on television audiences. The best case scenario for the audience researcher would be to find opinions about certain television programs, as this is information which cannot be submitted by traditional television audience research. For many television stations online opinions and communication on the internet about television programs has transformed into a kind of virtual currency (Wright, 2009) complementary to television ratings.

The common approach to finding communication data in the context of television is to search for key words or hashtags, using tools similar to search engines to find posts and comments by television viewers. Croatian television broadcasters have not started to promote communication about television programs on Twitter and they did not broadcast hashtags on air during the period this research was conducted. The question raised in this work was how to approach Croatian Twitter data with existing methods such as hashtag research (e.g. Bruns and Stieglitz, 2012; Bredl, et al., 2014), if Croatian Twitter users do not use hashtags to organise their communication.

Sampling was used as an approach to transferring traditional research elements to social media analytics. It will be researched if sampling submits a relevant number of key words, which can be used for further research (e.g. using opinion mining tools).

Finally, this work comes to the conclusion that traditional mass media is in charge of leading and building up its social media audiences. Without the strategic promotion of television programs and without the interaction between television broadcasters and their audiences, it seems that no or only limited communication in the context of television broadcast of the programs can be found. The worldwide trend is that (particularly younger) audiences prefer the internet and social media networks as a digital accompaniment to their lives (van Eimeren and Frees, 2009).

If traditional media wants to be part of this online world they need to develop a strategy to develop social media activity with the aim of building up a social media audience. Without that, users will still communicate on social networks such as Twitter about television programs, but finding this data will still remain a challenge. Television broadcasters are in the position to submit content to the (online) social media world, as they have the huge potential of submitting topics to talk about and to share. Multiple forms of television content distribution challenge television channels, as they have to evaluate their benefits and take into consideration that they have to protect their core business: ensuring advertising revenues. Social media networks are transforming towards becoming anew distributor of content. The television markets are in a process of change, but as the performance of television broadcasters shows, high quality television programs are still highly appreciated by television audiences. It is the challenge of the digital age to combine traditional media consumption and social media activities to follow the audiences on their way through the digital environment.

2 TECHNOLOGY CHANGES TELEVISION MEDIA

The media landscape and audience research have been transforming for a few years now. Additionally, over the years new digital technology has significantly changed and improved the television experience, with an increasing number of channels and television functionalities (Prior, 2007). "Associated with the traces that particularly the users of 'social media' (Facebook, Twitter, YouTube, etc.) leave behind and which constitute a wealth of both scientifically and commercially interesting evidence, big data are one outcome of a wide and deep process of digitalization – a process that is affecting not just how individuals, groups and entire societies communicate but also how research may be in a position to measure and interpret the consequences of their communications at local, national, regional and global levels of social organizations." (Jensen, 2014, p.227).

The share of internet users among the world's population is steadily increasing, albeit at a reduced speed. In Croatia, while one can speak of a complete internet supply for younger users through the increased use of mobile devices, a growing section of the population is gaining access to the internet and social networks thanks to the expansion of the digital network in rural areas. Television is affected by these changes in many aspects and empirical research on audiences is ever more important for audience analysis. Audiences become "less predictable, more fragmented or more variable in their engagement with media, understanding the audience is even more important for theories of social shaping, design, markets and diffusion than, perhaps was true for older media" (Livingstone, 1999, p. 4).

For the first time since the introduction of television, a change in television consumption worldwide can be observed (Nielsen, 2015; The Guardian, 2012). Despite the fact that the daily hours of usage remain high, younger people especially are turning away from traditional television habits and often use cellular devices (second screens) in parallel to their television consumption. There is a tendency for the viewer to divide their attention while using the cellular devices with its extensive communication possibilities in parallel to television consumption. Statista (2016c) reported that 69% of Croatians use their mobile phones to access online content while they watch television, 34% use computers and 11% use tablets.

Digital media, especially social networks, have been slowly changing how people communicate and the digital mobilisation in Croatia is growing and mostly interesting to an urban minority (Car and Andrijašević, 2012). Probably the most valuable development caused

by digitisation in Croatia is that information has become more available and censorship is no longer possible, one can speak of a real democratisation of news and once published the development of digital media has resulted in a pluralism of comments and voices.

The transformation of the television landscape and television usage go hand in hand with the change in audience research. Through the observed increased use of videos (YouTube) and the use of non-linear, i. e. time-delayed program offers, it will become more difficult in the future to reach the younger target audiences. Against this backdrop, it is of particular importance that Croatian television stations know their audiences and that the validity of traditional viewership rating research is challenged. Add to that a seemingly unlimited, previously unheard of volume of data whose usability appears attractive to the researcher, as in this research approach, viewership ratings of television programmes are also measurable.

This chapter seeks to present the changes to television media through digitisation and market fragmentation and to put the changes in the context of big data, created by "communicating audiences" (Jensen, 2014). Following this introduction, the chapter is organised into four parts: The first section considers the development of television with news kinds of media offers. The internet is given particular significance in section two, as the internet and its diverse communication methods which are available at any time through mobile devices are likely to have the largest effect on the development towards a social television landscape. Following that, the third section focuses on the special properties of the Croatian television market to contextualise and embed the observations, although the research is configured in a way which could be applied to other television markets under digital conditions. The final part of the chapter then shows how the changing television landscape and social web also have an impact on audience research.

2.1 Digitalisation of television

New technologies have always had an impact on television. In the last decade the digitalisation of media and the increasing internet usage and social media activity of television viewers have changed the world of television. The development of media and communication looks back on a history of changes driven by the technological progress of media and communication applications and its impact on the industry (Wirtz, 2011). The main milestones of technological media developments are presented in figure 2.1. This work focuses on the change of television and its (social) environment mainly influenced by the internet. Television

looks back on decades of stability where markets balanced public and commercial television offers and mainly technological developments could be observed. In particular the last decade brought, along with social internet and wireless environments, many innovations and changes that are more radical. “Television is now no longer what we used to assume, it is no longer what we felt to be ‘television’ well before we actually defined it.” (Pasquali, 2012, p. 1).

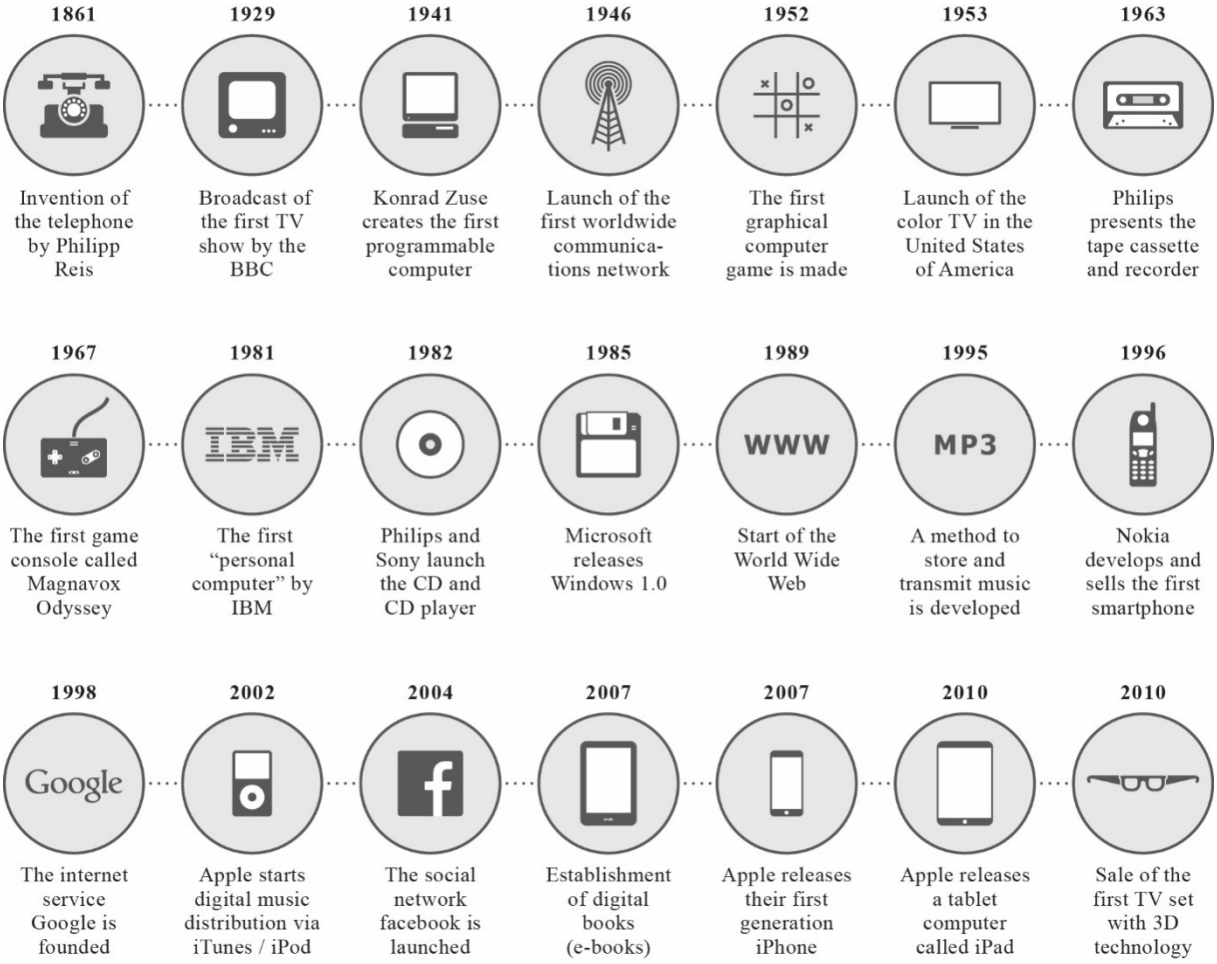


Figure 2.1. Milestones of television media development (Wirtz, 2011)

The first television show was broadcast in 1929 by the public broadcaster BBC, and in 1953 colour television was launched in the US. In 1989 the World Wide Web started and in 1996 Nokia developed and sold the first mobile phone. The social media platforms Facebook, YouTube and Twitter were founded in 2006. In 2006 television programs and films were

transmitted via internet for the first time, and in 2010 the first television set with 3D technology was sold. In 2010 the first iPad was on the market (Wikipedia 2016). The last decade was characterised by rich and rapid development, compared to previous decades (figure 2.1).

For the television industry the development of the internet was considered the basis of the new forms of television viewing as we know it today. The following paragraph portrays the development of the internet and its impact on the television industry. The internet is characterised by manifold functionalities. Simply describing the internet as a network does not convey the complexity of its meaning to society. The internet transformed technology and spheres of life. It provides technological infrastructure and creates public spheres of communication for societies, although the public sphere has never materialised "because of unequal access to communication channels, uneven distribution of communicative competence, and the reduction of public debates to a legitimisation of dominant opinions created by either the 'business type' or the 'government type' of power elites" (Splichal, 2007, p. 242). The following paragraph describes the functionalities of the internet and technological infrastructure (Züll, 2014) in the context of television.

2.2 The internet and social web

The internet is the worldwide interconnection of computers and electronic devices; it provides the technical infrastructure, connects people and enables users to exchange information.

The world`s main system of communication shapes our societies and daily routines as a virtual space in various ways: growing social media platforms and networked communication create tools with a new technological framework for those who had not previously been able to communicate to the public (Züll, 2014). "New platforms create openings for social, cultural, economic, legal, and political change and opportunities for diversity and democratisation for which it is worth fighting" (Jenkins, Ford, and J. Green citied in Züll, 2014, p. 120).

The possibilities of sharing an individual opinion with the public are expanded by the internet and the establishment of communication networks is enabled. Social networks are 'spaces of autonomy' and users are able to form networks and coordinate actions and social movements by connecting to each other, and they therefore create communities (Zittrain and Palfrey, citied in Züll, 2014). According to world internet statistics (2015), around 3 billion people worldwide had access to the internet at the end of 2015 and the numbers of participants

is still growing. The internet is considered to be the mass communication media of the digital age, connecting people to endless networks which transmit information all around the world (Castells, 2015).

In 2015 the average rate of internet penetration measured in terms of access was 46.4% (figure 2.2) of the worldwide population. This has been increasing over recent years and wireless mobile usage and technological convergence have been considered the driver of internet development (Castells, citied in Züll, 2014).

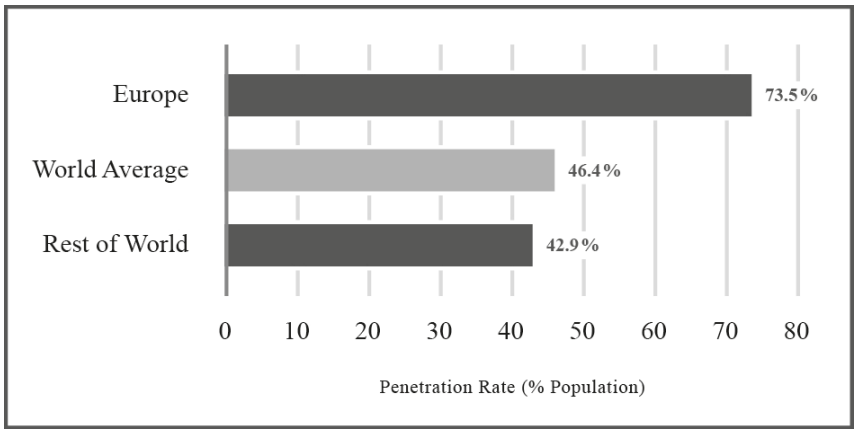


Figure 2.2. Internet penetration in Europe, November 2015 (Internet World Stats, 2016).

In Croatia, 3.1 million people used the internet. As Croatia has 4.2 million inhabitants (chapter 2.3.1), this is approximately 73.8% of the Croatian population and therefore, as can be seen in figure 2.2. Compared with European numbers this rate is close to the average measured in terms of access in Europe and far above the world average. A digital divide between the generations born before the internet age, and the generations who are growing up with the internet – so-called digital natives – can be observed.

The internet and expanding social media platforms have enabled various new forms of participatory activity. Virtuality has become an essential dimension in our lives and a shift from traditional mass media to a system of horizontal communication networks around social media has introduced a multiplicity of communication patterns (Castells, 2007).

According to a worldwide unique long-term study (1964-2010) about media usage and media evaluation in Germany (Reitze, Ridder, 2011), which is performed every five years on behalf of German public television, the main motivation of internet usage is communication: emails, instant messaging, chats, forums, and online communities (p. 64). The mass phenomenon of participation and communication via social networks confirms this global trend.

- Although the shared viewing experiences of television viewers has been undermined by a rise in the number of television stations, the growing multichannel environment, more television sets in homes, and the option for users to watch time-shifted viewing from DVRs, the internet acts against this fragmentation (Futurescape, 2012). It lets individuals discuss, recommend and share television content with each other, typically via social networks and social television applications. Second screen devices such as laptops, mobile phones, smartphones and tablets are ubiquitous in the home – and bring with them the opportunity to log in whenever and wherever users want. Second screens support the usage of other media platforms parallel to the viewing of television, and invite viewers to participate in social activities around television in real time. For example, viewers are according to Habajec (2014) invited to:

- use the internet as an enhanced television guide to access more information about a television program or advertising

- participate in interactive social television programs, e.g. via special apps and by voting in games and competition shows, alone or by engaging with others

- sign up for loyalty and rewards programmes of advertisers to receive virtual points or rewards for viewing engagement

- share any kind of information related to television such as content, photos and links related to the program

- share opinions with (virtual) friends who are having the same viewing experience, via texting, blogging, chatting, or tweeting

- participate in active networking (using real names or anonymously) in a digital environment

The invention of the internet and web 2.0 connected people and high internet penetration and growing social media platforms transferred communication to the online world. Watching television has always been a shared experience and social networks allow the audience to discuss what people are watching at any given moment, to share opinions and emotions and to start a conversation in a short and efficient manner (Bjur 2009; Wohn and Na, 2011). "People congregate in global virtual neighbourhoods such as discussion groups and chat rooms to engage in topics ranging from entertaining trivia to philosophical issues" (Sudweeks and Simoff 1999, p. 44). They spend their time on the internet at work or during their leisure time – browsing, searching, playing, texting, blogging, chatting, tweeting, and networking. Research has shown that people enjoy watching television as a part of socialisation in groups (Duchenaut, et al., 2008).

2.2.1 The internet as the source of big data created by television audiences

The growing popularity of the internet, and the internet itself have attracted "the attention of a vast body of researchers, ranging from philosophers to technocrats" (Mitra and Cohen, 1999, p. 180). "The metaphor of the internet as a market-driven social space lends itself particularly well to market research that has long desired predictive precision at the level of the individual consumer and has employed a variety of technologies with which to gather sufficient information in an attempt to ensure predictive power" (Jones 1999, p. 4.).

Whenever people interact, data is created, and this is given a great deal of attention in internet research (Jones 1999). Audience researchers are constantly exposed to new sources of data, leading to the issue of how to evaluate the opportunities and challenges that the data presents, in order to study television audiences. Social media services function as exchange points for information, users leave digital traces of their actions, and big data crosses the internet, is stored, and therefore becomes accessible to researchers to analyse.

The enormous potential of the knowledge resulting from digital communication on the internet is used in a number of ways. Google, Wikipedia and other similar websites have been built on the diverse knowledge of amateurs, and big data collections can be gathered by the masses more accurately and quickly than that of individual experts (McAfee and Brynjolfsson, 2012; Mayer-Schönberger and Cukier, 2013). This principle is known as crowdsourcing or swarm intelligence. The complex project of asking masses of television viewers their opinion

about a television show or channel is broken down and performed by many people subconsciously, and the crowd comments in parallel. The analysis of the social response of the masses is the main aspect of social media research, which is similar to the concept of swarm intelligence and is different to the concept of the wisdom of crowd, which is based on the concept that the many (seeking the average) are smarter than the few.

The exclusion of individuals from the internet either because they have no access to the internet or because they do not want to use the internet influences the representativeness of data analysed by the researcher when using the internet as tool of market intelligence. The usage of the internet and in particular of social media tools by television audiences is a precondition for tracking audience data. Social media providers are masters of this data, scientists try to find patterns in data and translate them into useful information for the industry, and whole organisations need to redefine the evidence that using big data intelligently could improve business performance and will transform how we live (McAfee and Brynjolfsson, 2012; Mayer-Schönberger and Cukier, 2013). Search engines such as Google and Yahoo and social networks store every query, cookie and log-in data, keep records for many months, and create big data warehouses. Keeping individualised records stirs the interest of researchers, but the awareness of end users must also increase. As electronic transmission may be unsecure, this data may provide insights into an individual's personality, habits and interests.

The aim and philosophy of social networks is to organise the information and communication data of the world and to make it accessible and usable to everyone. This mission becomes problematic in countries with authoritarian governments controlling the internet (Züll, 2014). The social media platforms as huge conglomerates commercialise their data by selling it to the industry – such as television – which is interested in reaching the consumer by targeted advertising. The volume of comments can be measured and Züll and Mikelić Preradović (2013) asked "if a correlation between TV ratings and social media activity" exists. The technical infrastructure of the internet integrating an archive of user data offers new opportunities for audience researchers. Jensen (2011) portrays internet research as a tool for and object of analysis as the research can produce data and also find data.

The tracking and usage of internet data and the question of access to this data - especially in relation to user identification – raises questions about privacy and liability. The aim should be to carefully create and define balance between research and user privacy, particularly if research data is used by the private sector industry or public institutions.

The protection of the privacy of government and industry is given high priority in democratic nations and users expect their data to be used responsibly (Züll, 2013).

2.2.2 With digitalisation towards fragmentation of television

By the end of 2011 all Croatian television broadcasters had switched to digital transmission (Car and Andrijašević, 2012). The digitalisation and fragmentation paved the way for new channels and television services. Furthermore, television delivers all digital formats to their audience on second screen devices such tablets, laptops, mobile phones, game consoles and interconnected televisions. The digitalisation has demanded that the government and industry invest in digital infrastructure, new services, and the production of programs and metadata.

Thanks to this transition, television will change more in future decades than it did in previous decades. And due to the increased space for transmission of television programs, broadcasters‘ distribution costs of the television signal decrease and the transmission capacity limit is overcome. The increase of digital television channels can be observed worldwide. The media markets have been more and more fragmented and the usage of media has become more and more individualised.

Today, people watch television at home and on mobile devices whenever and wherever they want. Audiences use second devices (second screens) while watching television and are increasingly engaged in different forms of social participation. Different angles of the same story are brought to these audiences through different screens: computer, television and cinema screens, mobile phone displays, portable MP3 players and video player displays (Hesmondhalgh, 2007).

With digitisation, a growing number of television channels compete for the attention of its audience, and therefore new distribution channels for television content are created. The major television networks have founded niche channels to better reach selected target groups which are important for advertisers. IPTV (internet protocol television) enables television channels to transmit and spread its television programs via the internet.

With an increasing number of channels, there are more thematic channels and can be integrated into the packages of multichannel platforms. Thus, the internet can be used purely as a transmission channel for television. In the USA, young people have chosen to live without a television set and the development of new platforms offering time shifted television viewing is

currently affecting trust in traditional television audience measurement. In particular, the fact that digital video recorders (DVR) now allow the audience to skip commercials has led to uncertainty in markets with a high usage of digital video recorders and broadband internet access.

Television viewers are able to watch programs whenever and wherever they want on a number of electronic devices. Digitisation has been the prerequisite of media convergence, the term which covers "the flow of content across multiple media platforms, the cooperation between multiple media industries, and the migratory behaviour of media audiences who will go almost anywhere in search of the kinds of entertainment experiences they want" (Jenkins 2006, p.2). Since people have started to consume different media at the same time, media convergence seems to be the solution for the television industry to the problem of keeping the attention of television audiences.

The leisure time for media consumption is limited and needs to be allocated by the media consumer among various media offers such as television, radio, internet and newspapers/magazines (print). Media usage in Croatia is still high and most Croats watch television daily. With 17.5 hours of television consumption per week figure 2.3 shows that television is still the most consumed media in Croatia.

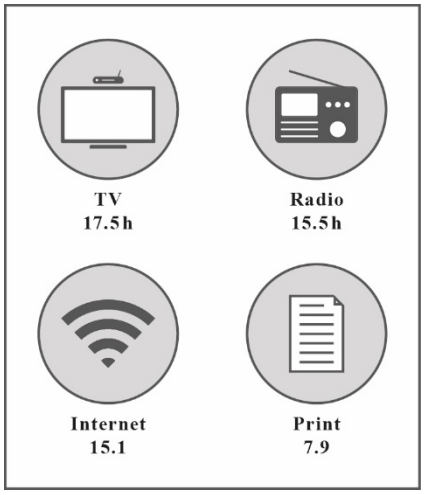


Figure 2.3. Media usage in Croatia (Konrad Adenauer Stiftung, 2016).

Although internet usage is growing, television is the main platform for information, education and entertainment in Croatia. The amount of time watching television has dominated

the routines of daily media usage for a long time, and one can argue that it dominates what we think (Livingstone, 1998a, p.4) and what we communicate and share. Therefore, television remains – like in many other countries of the world – the most used medium for information and entertainment.

The average media usage, which includes television, radio, newspapers and the internet, reaches a maximum of approximately 10 hours per day (Best and Breunig 2011, p.16), and people increasingly use different media in parallel.

A relatively high amount of television is still being consumed in Croatia, in comparison to other European countries. Although it can be seen that daily television consumption has, according to figure 2.4, decreased slightly, television consumption is stable, at about 4 hours a day, despite fragmentation of television markets.

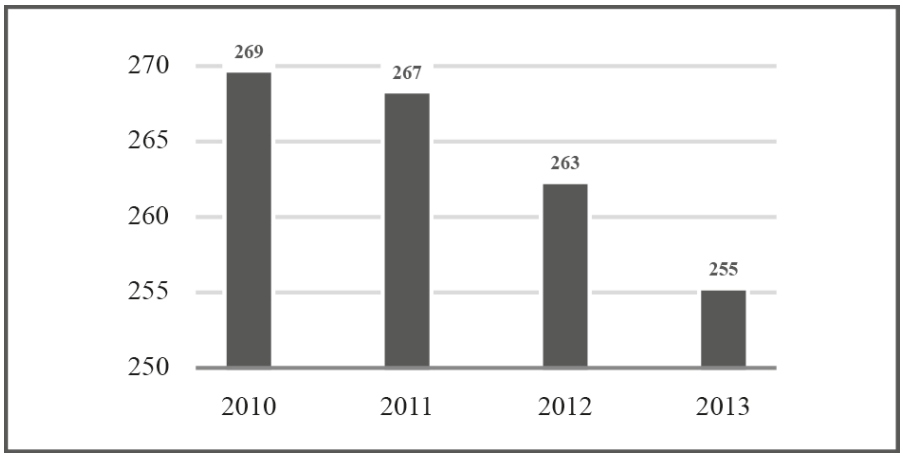


Figure 2.4. Levels of viewing time of individuals in Croatia aged 4+ (AGB Nielsen Media Research cited in RTL Group, 2013).

The television industry has been faced with the situation that television viewers split their attention among several electronic devices and that they use media in parallel. The use of tablets, laptops and/or smartphones accompanies the consumption of television; this is often described with the term *second screen*. The parallel usage of media increases if different media are compatible and can be used complementarily.

If media target different reception channels such as television and internet, the probability of compatibility increases. Mobile internet and social networks mainly attract younger people – the so-called "digital natives" – who grew up with the internet, as opposed to the 'digital immigrants' who had a certain period in life without the internet and digital media. Best and Breunig evaluated socio-demographic characteristics of parallel media usage and came to the conclusion that multitasking and the parallel usage of media varies according to age and level of education. The parallel usage of different media seems to be typical for younger, educated people. However, watching television is a communicative and social activity.

For television channels this means that they partly lose the attention of their viewers, but at the same time they gain the chance to interact with television viewers via special second screen applications. Together with the proliferation of mobile second screen devices it provides virtuality for a connected television audience that uses social media platforms while watching television. As a consequence, television does not command the full attention of television audiences (figure 2.5).

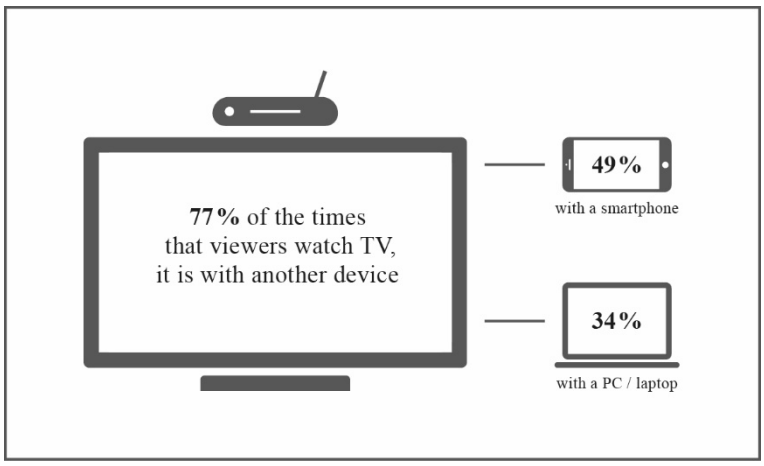


Figure 2.5. Social television and second-screen viewing (the Guardian, 2012).

Media convergence is the merging of previously individual media and the merging of information and communication technology because of digitisation. Media convergence leads to changing audience behaviour in a modified media environment as the media convergence leads to a diversification of media offers.

All Croatian television broadcasters have an online presence offering branded pages of information and entertainment. Television channels use their websites mainly to promote and provide information about their programs and they usually launch branded program websites. “Impetus for developing new digital applications is given by the emerging trend of second screen consumption, which means the usage of other devices like smartphones, tablets and/or laptops while watching TV” (cited in Züll, Mikelić Preradović, Boras p. 278, 2013).

Content providers such as HRT, RTL Hrvatska, and Nova TV develop second screen applications to build up connected television experiences for their audiences. Social media activities of Croatian television channels are diverse and developing. Television related interactive tools and applications which are designed to promote social interaction and participation are offered to the audience. For example, RTL Televizija introduced social viewing on its video on demand platform RTL Sada (2016). It enables the viewers to talk and share in real time what they are watching and what their opinion is about the program being watched. RTL Televizija offers an application for second screens, where people connect RTL Sada with their Facebook account so that Facebook synchronises the social watching activities and enables Facebook friends to follow each other’s activities (figure 2.6).

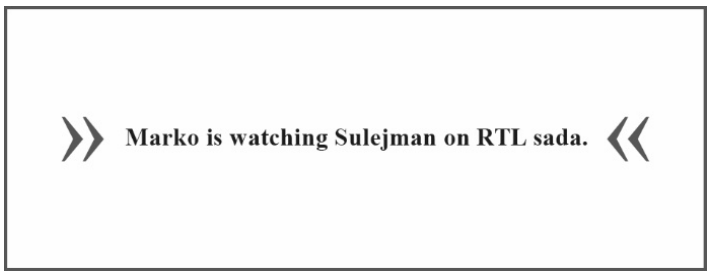


Figure 2.6. Quote on RTL sada (sada.rtl.hr, 2016).

The research questions investigated in studying mass media are virtually unlimited (Wimmer and Dominick, 2011) and this field of research does not lack predictions and visions about the consequences of media convergence (Reitze and Ridder, 2011).

People started to use mobile devices to watch online television programs, young people started to live without a television set and the development of new platforms offering time shifted television viewing is currently affecting the trust in traditional television audience measurement. Due to the new media usage patterns, there seems to be a need for more insights

into television audience behaviour. Traditional television audience research has not been able to provide the required insights, given the limited data situation. The new social habits of television audiences create a theoretical conception of a more active audience (Livingstone, 2005), an audience which communicates and shares on social media platforms.

2.2.3 Social media and television

The expansion of social media platforms and their integration into our daily life is a worldwide phenomenon. There is a general trend of social media transcending borders, age barriers, and cultural barriers, and social media actually becomes a driving force of transformation of cities within a digital environment (Züll, 2013).

Different platforms require different kinds of engagement of users, and various forms of content and sharing characterise different social media platforms. Content communities such as YouTube and social networks such as Facebook ask the user to disclose information in the form of personal user profiles and other personal data. Social media platforms are often privately owned and social media data is controlled and used by social media platform providers for commercial purposes. Procter, Voss and Lvov (2015) cited Savage and Burrows who, in the increased availability to the private sector and a drive to analyse it, identified the "Coming Crisis of Empirical Sociology".

Different definitions for social media can be found in literature (Cohen, H., 2011; Lake, C. 2009; Kaplan and Haenlein cited in Züll, Mikelić Preradović, Boras, 2013), in general describing internet-based tools, where users can be interactive and participate by creating and sharing different kinds of content.

Social media has made new forms of communication and connection among individuals possible and it enables individuals to express themselves, to extend personal reach via virtual spaces and supports the free flow and exchange of information. Kaplan and Haenlein (2010) define *social media* as "a group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" (p.61). They submit a classification for social media (table 2.1) and put them into different groups, where the distinction is made based on the "degree of self-disclosure it requires and the type of self-presentation it allows" (p. 62).

Social media platforms offer their users the opportunity to collaborate and to share and exchange content in virtual communities. Users of social media networks can create content together (Wikipedia, blogs, Twitter) in collaborative or individual platforms, they can share content such as videos and photos (YouTube, Flickr; Instagram, Pinterest), or take part in social networks, find friends, follow and communicate with users wherever they are (Facebook) and participate in games and virtual worlds.

Table 2.1. Classification of social media (Haenlein and Kaplan, 2010, p. 62).

		Social presence / Media richness		
		Low	Medium	High
Self-presentation / Self-disclosure	High	Blogs	Social networking sites (e.g. Facebook)	Virtual social worlds (e.g., Second Life)
	Low	Collaborative projects (e.g., Wikipedia)	Content communities (e.g., YouTube)	Virtual game worlds (e.g., World of Warcraft)

Mass media join social networks and participate as *users* to generate and distribute content. Social media platforms as networks on the internet continue to grow enormously, reflecting society's desire to network and to share. Social networks are expanding in size and popularity and are becoming a more prominent part of our everyday life. Currently more than 1.6 billion social network users (Statista, 2014) access social media services online and more than half of internet users are also social network users.

Despite Facebook having 1.35 billion users as of December 2014 (figure 2.6), making it the most popular social network worldwide, other social networks have been growing heavily and the spread of smartphones and tablets furthermore support this trend. As of January 2014, 71% of online adults use Facebook (Pew Research Centre, 2013). The growing ubiquity of smartphones further increases social network usage.

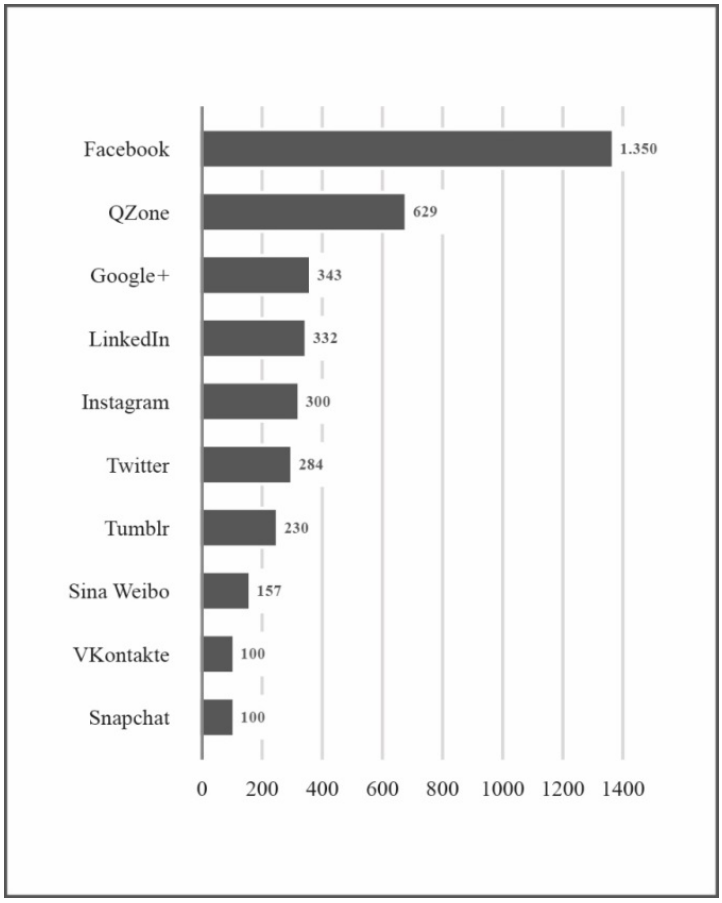


Figure 2.6. Leading social networks worldwide (December 2014), ranked by number of active users in millions (Statista, 2015a).

The high usage numbers of Facebook could first lead to the conclusion that this would be the platform to use for research. However, the quality of analysis depends heavily on the openness of the provider in respect to methodology and transparent data sources. Facebook is restrictive regarding its data policy, claiming to protect the privacy of their users. The Facebook privacy settings affect the type of data which can be accessed by the researcher (Giglietto, Rossi and Bennato, 2012).

With the strong commercial interest of Facebook in user data which is inherent with their protection of privacy (Liu, et al., 2011) it has become apparent that studies on Facebook data are rare. Twitter is considered more effective at live coverage than Facebook as the usage of short messages and hashtags (#) submits quickly and has an immediate reach (Hightfield, Harrington and Bruns, 2013b). A hashtag is a type of label used on Twitter or other social media networks which makes it easier for users to find messages with a specific topic. Interesting research concerning streaming analysis can be found if Facebook participates and supports the

study by giving access to its data, as reported by Giglietto, Rossi and Bennato (2013) referring to an agreement between the Politico website and Facebook, which focused on the 2012 U.S. Republican primary elections and opinion mining of the comments.

Some social networks cater to huge local audiences, such as Qzone in China with 629 million users, while others specialise in particular topics, such as the professional network LinkedIn with 332 million users. Instagram (300 million) and Twitter (284 million) are next, followed by Tumblr, Sina Weibo, Vkontakte and Snapchat (figure 2.6).

There is a general trend of social media transcending borders, age barriers and cultural barriers and social media actually become driving forces of transformation. The high usage numbers show that social media is more than a phenomenon – "it is the transformation of a society in a digital environment" (Züll, 2013, p. 211).

.Social media allows people to share and disclose content in multiple ways, such as uploading videos and photos, forwarding links and in particular communication via text, often enriched with emoticons which contain the exchange of information and opinions towards any kind of topic. "A spreadable mentality focuses on creating media texts that various audiences may circulate for different purposes" (Jenkins, Ford and Green, 2013, p. 2).

Sharing photos is the most popular activity on social networks and 26% like to voice their opinion towards a topic (figure 2.7).

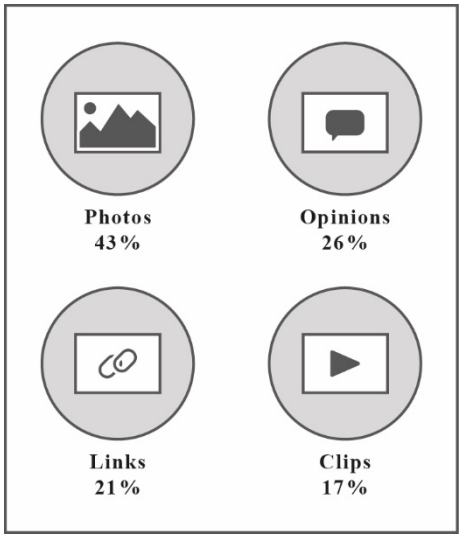


Figure 2.7. What people share on social networks (Bennett, 2014).

In general, the television industry produces and distributes television content in different formats and for different platforms and is "as mass media practically predestined to produce content which can be discussed and shared on social networks" (Züll and Mikelić Preradović 2013, p. 124). However, digital communication technologies and the proliferation of digital communication media such as social networks support the participation of television viewers in a collective process of media consumption and have generated connected audiences that communicate amongst themselves about television content. "Such online discussions provide a new means to sense the public interests and generate feedback in real-time, and are mostly appealing compared to generic media, such as radio or television broadcasting" (Wu, Zhu, Wu and Ding, 2014, p. 98).

Television channels started to produce different forms of texts, not only television programmes, but content for their online presence. With growing social media and increased participation of audiences in social networks, television particularly adapted to young audiences, building and maintaining relationships via social media. They established a new presence in social media with Facebook pages, Twitter accounts, and YouTube channels, and designed new forms of television-related content such as pictures, texts (editorial or user generated), videos (editorial or user generated, short or long, music video or video), and links.

However, media convergence reshaped the relationship between television audiences and broadcasters, and today content can be found across multiple platforms (BBC, 2011). With the use of digital technology and the rise of interactive functionalities, the new kind of social text is more individualised (Danesi, 2002). With the introduction of the internet and social media, the traditional passive audience changed into an active audience in terms of both the content and the agencies producing that content. Social media and internet made it possible for everyone to talk and write about everything and publish it online (Lessig, 2005; Castells, 2007). "The audience has turned into audiences, and the consumer can always and easily become a producer-author. At present, there are blog authors with bigger daily audiences than national televisions or newspapers" (Chiribuca, Pah and Hunyadi, 2008).

Social media facilitate communication around particular media events or topics of interest (Highfield, 2012). Highfield, Kirchhoff and Nicolai call this phenomenon 'topical networks' which "refers to the collection of sites commenting on a particular event or issue, and the link between them" (2011, p. 341).

Television programs are promoted by using social media campaigns. The data produced by television viewers is of high interest to the television industry as it helps to figure out what people want and can therefore be valuable to networks and advertisers (Arndt, 2011, Edelsburg, 2011, Friedman, 2012, Friedman, 2009).

Increasing mobility and interconnectivity characterise the shift towards an *active* participatory media spectatorship (Jenkins, 2010). Over decades television consumption has been passive, the television viewer was alone or with family and the television viewing took place in a *one*-screen world and with a limited number of television channels. Television consumption is now more personal and individualised, but also more social than ever (Pasquali, 2012). Due to digital change of the media environment, television viewers consume media content on first and second screens, at home or when mobile, synchronous or asynchronous (Jensen, 2010) and moreover the research on content consumption on the web reveals a strong concentration of collective attention on a few items (Huberman cited in Züll, Boras and Mikelić Preradović (2013) within a short time-span (Leskovec cited in Züll, Boras and Mikelić Preradović (2013).

Traditional media act and react, they are an integral part of the changing media landscape. Internet and social media provide a non-hierarchical and multidirectional communication platform for journalists, web-based interest groups and individuals to explore topics ignored by the mainstream broadcast media (Plaesu and Dalu, 2008). Furthermore, most social media users want to decide about the relevance of an event or news on their own and dislike the notion of traditional media filtering the news for them (Beheshti-Kashi and Makki, 2013). Today, a networked television model can combine internet broadcast and streaming, SMS, and teletext (Cardoso, 2006).

Although the content is different and diversified, the broadcasting model seems to stay the same, as television broadcasters submit content to be shared and talked about. Social networks such as Twitter and Facebook seem to be developing more and more into broadcast mediums, while traditional media participate in social media as content providers and in the form of one-way-communication. It seems that the *broadcastification* of social media can be observed, since a small number of followers follow a large number of brands and *mega-publishers* (Levine, 2013).

Television channels have a presence on all important social media platforms with many Facebook pages, YouTube channels and Twitter accounts but they also provide social

offers within their own online platform environment, where people are invited to comment on something or to watch a video. Participating in social media networks means the creation of a huge number of websites and platforms which need to be managed.

What are the activities of television audiences and what is the overall benefit of the broadcaster engaging in social media? The positioning of television broadcasters and their audiences in the social media world raises many questions. According to Pasquali (cited in Züll, Mikelić Preradović, Boras 2013, p. 276), "there are at least 4 ways in which the traditional broadcast media intersects with social media:

1) real time TV viewing multitasked and supported with social media conversations;

2) access to mainstream TV content through different institutionalised and branded platforms, such as official websites, social networks profiles and forums;

3) mainstream and niche TV content grassroots downloading, sharing and archiving;

4) following TV content on YouTube or on other video sharing services.".

YouTube is the most successful video sharing website worldwide, not just because of broadband allowing streaming video but also because of the role that the visual plays within computer mediated communication (Howell, 2005). The intersection of the traditional broadcast media with social media creates a set of mixed styles of television consumption that design repertoires (Hasebrink and Popp, 2006) or media matrices (Cardoso, 2008) which enable the fusion of interpersonal and mass communication – connecting audiences, broadcasters and publishers and giving newly mediated roles to their users.

Digital technologies have increased the number of television channels and developed into a highly diverse broadcasting system with changing delivery platforms. The participation of mainstream media in social media has transformed television, and new digitised products for different social media platforms had to be created. This was the beginning of interaction with the audience, where the audience participates and television broadcasters try to maintain a relationship by creating individualised program offers in order to form a social relationship with the audience.

"The Social TV project builds on the increasing integration of television and computer technology to support sociable, computer-mediated group viewing experience" (Ducheneaut, et al., 2008, p. 136).

In social television the audience transforms towards a powerful ally of television. In general, *social television* describes internet-based digital interactions regarding television, or digital interaction with content by television viewers and its television channel. In the next chapter the Croatian television market will be described to contextualise the analysis of social media activities in Croatia. Television becomes socially interactive when channels create shows to build relationships with viewers by providing (social) interaction around television shows. In particular, big international competitions and real time live programs and television formats are appropriate for creating social television programs, where people share their television viewing experience.

2.3 Croatian television in transformation

The Republic of Croatia is a South-Eastern European country with a population of 4.29 million. From socialist times to the current democracy, the media in Croatia has been transformed through a series of reforms (Goldstein, 2010) towards the transition to democracy (Čular, 2000). The breakup of Yugoslavia and Croatia gaining independence in 1991 started the trend of market liberalisation. The idea of democratisation took hold rapidly and resulted in elections in 1990 (Kasapović, 1999). According to the Bertelsmann index of transformation report (Bertelsmann Stiftung, 2008), the democratic institutions in Croatia are stable, legitimised, functional and are accepted by all political parties. With steadily increasing development of television media and early digitisation over recent years, Croatia is a reference market for all of Southeast Europe.

Joining the European Union in July 2013, the Croatian government established a stronger democratic and regulatory environment to ensure the independence of the media from political and economic influence. The Agency for Electronic Media of the Republic of Croatia implements the procedure for granting concessions for the provision of television and radio services (2016), and today Croatia is characterised by a pluralistic and diversified television landscape with strong public and commercial free-to-air television channels. Different studies have been published which provide detailed information about the Croatian media landscape as

they monitor the media, and among many other responsibilities, they monitor their legal and constitutional framework.

Peruško (2011) conducted research for the United Nations Educational, Scientific and Cultural Organisation (UNESCO 2011) into the state of the development of the Croatian media system. Freedom House, a United States (US) based non-government organisation which conducts research and advocacy, investigates freedom of press in a country (2015). According to Freedom Press, Croatia is considered to be *partly free* (Freedom House, 2015). Even though the Croatian constitution recognises freedom of the press, there are a number of legislative restrictions on press freedom, and the state has tolerated harassment of journalists and nebulous ownership situations, in particular in the journalistic sector. Car and Andrijašević conducted research supported by the open society foundations, with the aim of mapping the digital media landscape in Croatia (Car and Andrijašević, 2012).

Free-to-air television (Free-TV) describes television services in unencrypted form but digitally encoded; it enables any television household with appropriate digital equipment to receive the signal and to watch the content without requiring any subscription or other ongoing costs or commercial fees, even though they may be delivered by another carrier. Some of the biggest carriers are Hrvatski Telekom which is now fully consolidated into T-Com/T-Home, as well as B.net or Vip.net, for which it is necessary to subscribe via satellite, cable or the internet. In a traditional sense, free-to-air television is transmitted by signals, which can be received via antenna.

The Croatian Bureau of Statistics is the main producer, disseminator and coordinator of the Official Statistical System of Croatia and main representative of the national statistical system for European and international bodies. It collects and processes many types of data, such as demographic numbers for the Republic of Croatia and also conducts censuses, the procedure of systematically acquiring and recording information of a given population. The most recent census was carried out in April 2011.

2.3.1 Characteristics of the Croatian population and general media indicators

Since 2007 the collection of media and television data, in particular the calculating of television ratings and television market shares, has been conducted by Nielsen, a global information and measurement company (Bjur, 2009) which is active in over one hundred countries. Among other activities, Nielson submits data to the television industry. The main

international standard system for producing television audience data is a system called the people-meter-system, which will be explained in more detail in Chapter 3. This applies to terrestrial, cable, and satellite free-to-air television broadcasters. Table 2.2 shows the broad national coverage of television in Croatia, where the average television households (TVHH) has three members.

Table 2.2. Country data of Croatia (Croatian Bureau of Statistics, 2011).

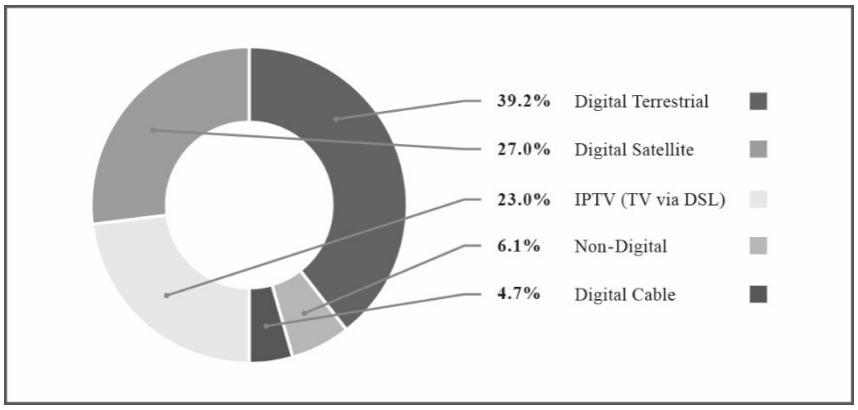
Country Data	
Land Area	56.594 km²
No. of Inhabitants	4.2 Million
No. of Households	1.5 Million
Total No. of TV Households	1.44 Million
Average Household Size	3.0

By the end of 2011 all Croatian Television households had switched from analogue signal distribution to digital television transmission (Car and Andrijaševič, 2012).

An advantage of digitisation is not only a better quality of television image compared to analogue television, it is also the liberation of space, as the signals are more compressed, therefore allowing for more television channels to be transmitted. Digitisation enables new forms of communication practices and content consumption on different forms of (mobile) devices as they can supply content quickly, at any time and place. The downside of the increasing number of information and entertainment channels may be the fragmentation of audiences which could significantly affect the television industry over the next few years.

Most Croatian television households (TVHH) are already digital, most households receive television via terrestrial distribution, transmission via satellite and IP-TV; cable television plays a minor role as can be seen in table 2.3.

Table 2.3. Digital television in Croatia, 2013 (IHS Screen Digest, RTL Group, 2014).



As nearly all households in Croatia have one or more television sets, television is still the predominant source of information and entertainment, even though the internet is a competitive source of information and news. However, for breaking news situations in particular, (extraordinary events such as natural or human catastrophes when channels interrupt their programs to report on these topics of high interest for the whole population), television is considered to be an important source of information (Lovreček, 2014). According to a study in Croatia conducted by Victoria Car (2010), the most trusted information platform by citizens is the internet (45 percent), followed by radio (39 percent) and television (35 percent).

2.3.2 Television landscape in Croatia

Historically the two television channels HTV1 and HTV2 operated by state-owned Hrvatska Radio-Televizija (HRT) are the most important as they have been the largest and most influential television stations in Croatia.

Both public channels were launched in 1956 (HRT, 2009) and HRT was considered in former Yugoslavia to be one of the main television stations (Terzis, 2007). Public television is mainly financed by a license fee and also by advertising, and is rarely funded by the state.

Table 2.4. General data of Croatia (Croatian Bureau of Statistics, 2011, Nielsen, 2012).

TV - General Data	
Total number of TV households (TVHH)	1,440,443
In % of all households	94.8
Multisets (homes with more than 1 TV set)	38.9
TV licence-fee cost per year	960.00 HRK 125.88 EUR

Commercial television finances its operation predominantly through advertising revenues, generated by selling advertising space to clients from the consumer goods industry, retail, and telecommunications. When broadcasting advertising, the advertising clients purchase *reach* which means that a certain number of television viewers can be addressed by the advertisement.

Looking at table 2.5 and 2.6, the channels HTV1, HTV2, Nova TV, and RTL Televizija are considered state-wide television broadcasters, given their television licenses and terrestrial distribution via the best digital multiplexes bringing them nearly full coverage over the whole territory of Croatia.

Table 2.5. Domestic public television landscape in 2013. (RTL Group, 2013)

Domestic Public Channels

Channel	Launch	Tech. Pen.	Programming	Revenue	Sales House
HTV 1	1956	99.2%	Generalist	L-F, Adv.	Hrvatska radiotelevizija
HTV 2	1956	99.1%	Generalist	L-F, Adv.	Hrvatska radiotelevizija
HTV 3	2012	92.0%	Documentaries, Film	L-F	None
HTV 4	2012	n.a.	News	L-F	None

Table 2.6. Domestic private television landscape in 2013. (RTL Group, 2013).

Domestic Private Channels / Foreign Channels With Local Ad Window

Channel	Launch	Tech. Pen.	Programming	Revenue	Sales House
RTL Televizija	2004	99.4%	Generalist	Adv.	RTL
RTL 2	2011	95.3%	Entertainment	Adv.	RTL
Nova TV	2000	99.4%	Generalist	Adv.	Nova
Doma TV	2011	94.5%	Entertainment	Adv.	Nova
Sportska TV	2011	77.9%	Sports	Gov., Adv.	In-house
Croatian Music Channel	2011	83.4%	Music	Adv.	In-house

The dual broadcasting system, which means that there are public and commercial television channels in one media system was introduced in the 1990s (Peruško, 2011). It is characterised by a strong state-owned public broadcasting group (HRT) (initially with two channels), (above 99% penetration) and two commercial television broadcasting groups, Nova TV and RTL Televizija.

During recent decades the HRT channels were subject to close political control, particularly until the 2000 government was elected. However, the evening HRT television news program was still seen as the predominant source of information by the Croatian public (Lovreček, 2014).

In 2000 the television market was opened by the government and Nova TV was founded, the first commercial television station in Croatia with full national coverage, today owned by Central European Enterprises (CME), a leading media company in central and Eastern Europe. CME again is controlled by Turner Broadcasting which is owned by Time Warner.

In 2004 RTL Televizija was launched by Radio Television Luxembourg (RTL), the biggest free-to air media broadcasting group in Europe with fifty-seven television stations and thirty-one radio stations. Each day, millions of viewers all over Europe watch RTL Group's television channels regularly, the biggest television group in respect to reach and advertising revenues is RTL television with its channels in Germany. RTL Televizija is also a commercial channel with national coverage in Croatia.

At that time HRT was not prepared for competition: Croatian audiences liked the entertainment approach of the new commercial channels, which was already dominated by big

US feature films, international television series, and sitcoms and they started to build up their audiences with increasing market shares and television ratings. HRT had to adapt their program schedules to keep their viewers and had to develop more modern and entertaining image while keeping the principles of public television in mind, which include having universal appeal (appealing to general tastes and interests), devoting attention to minorities, contributing to a sense of national identity, and promoting languages, dialects, and community. Public television must broadcast determined shares of informative, educational, cultural and entertainment content in order to fulfil their public mandate.

By media legislation, Nova TV and RTL are obliged to produce and broadcast daily news and informative programs in prime time to support a democratic and pluralistic media system.

With the analogue switch-off at the end of 2011 and the start of digital terrestrial transmission (DVB-T), in addition to these four main national television channels, another two channels were launched by Nova TV (Doma TV) and RTL (RTL2) which inhabit the second channel' position of the main channels. They are mainly used for the broadcast of so-called library programs, which are programs which have already been broadcasted on the main channels. Similarly, HTV was given two further television stations, HTV3 and HTV4, which are not allowed to contain any advertising material. In addition, another twenty commercial channels were launched on a regional level with limited coverage; this limited reach for advertisers means ongoing problems as financial sources are limited.

2.3.3 Advertising as the main financial source for commercial television

Like in many other European countries, public television in Croatia is mainly financed by a license fee, which must be paid by any household which is equipped with a television and radio set. In addition, public television stations in Europe compete on advertising revenues although they are restricted in respect to certain periods of the day when they are, for instance, allowed by European Union Law (eu-lex.europa.eu) to broadcast commercials and there are restrictions in respect to prime time (broadcasting time after 8 pm) and the number of commercial breaks permitted.

Overall, television (commercial and public) earns the biggest share of media spending. The timeline in figure 2.7 displays the value of media expenditure by advertisement and type in Croatia from 2007 to 2015, with a forecast at that time for 2015. Advertising spending on the

internet was expected to increase and reached 150 million Croatian Kuna in 2015. The share of media expenditure on television decreased over recent years to 700 million Kuna. The loss of media expenditure on television is caused by the development of the economy as the industry has to lower media budgets during periods of lower economic prosperity.

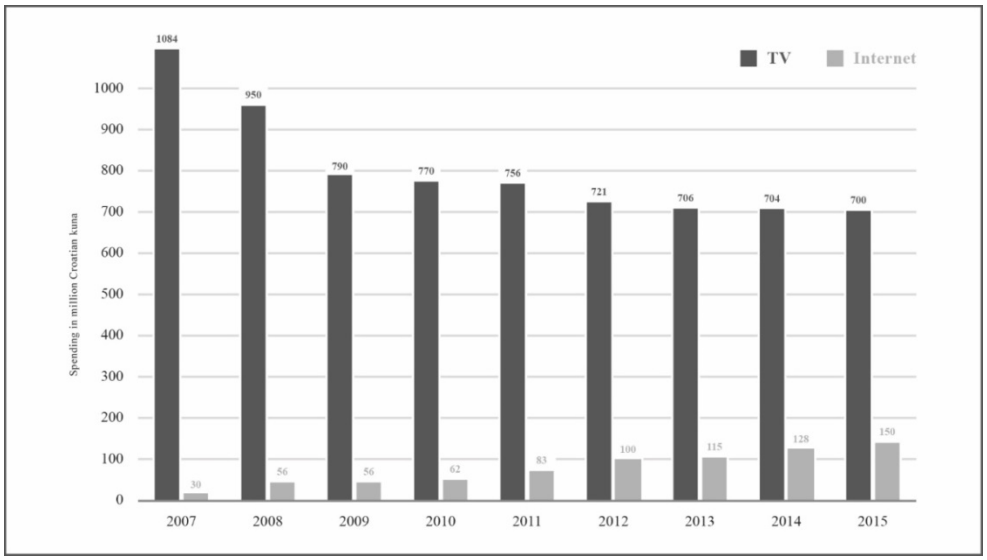


Figure 2.8. Spending on advertising in Croatia 2007 to 2015, in millions of Kuna (Statista, 2016b).

2.4 Audience research in transformation

Taking the audience research and measurement technologies as the starting point of this research work, the changes in television and media motivate to face the challenges that audience research has to face in the upcoming years.

In light of the ongoing technological developments the oldest and widest intergovernmental network for cooperation in research COST (European Cooperation in Sciences and Technology), established by the Ministerial Conference in 1971, started COST Action ISO906 *Transforming Audiences, Transforming Societies (2010-2014)* to coordinate research efforts "into the key transformations of European audiences within a changing media and communication environment, identifying their complex interrelationships, with the social, cultural and political areas of European societies" (2010-2014). Changing audience practices created by ongoing sociotechnological developments created a need for appropriating research methodologies (Patriarche, et al., 2014) and the contribution of researchers to COST reflects the need for creativity in audience research and illustrates the "dialogue between research traditions". However, bridges between academic and commercial research must be built as not a great deal of data and references about "transforming audiences" can be found in Croatia yet.

The research group COST presented, with one hundred and thirty-three scholars attending the conferences of this group (one was held at the University of Zagreb, Croatia on April 7-9, 2011), the methodological agenda of today's audience research in a changing media environment. As the author participated in COST activities this work contributes to the development and application of microblogging as an approach to Croatian audience research studying the conversations referring to television broadcast in Croatia and the relevance of instruments analysing Twitter content.

Television channels and the advertising industry target television audiences with their program and advertising content, to build up network and brand loyalty. Television audience research submits data to the market participants to enable them to act based on commonly agreed to objective methodologies and common data to measure success. Television created "the largest 'imaginated community' the world has ever seen (the TV audience)" (Fiske and Hartley, 2003, p. XVI) and audience measurement in general helps to understand the behaviour of people watching television.

Audiences are essential to the operation of television and all other mass media. The word audience in the context of mass media is traditionally used as the collective term for the "receivers" of the message (McQuail, 2000, p. 360).

As outlined by Carpentier, Schrøder and Hallett and referred to Jensen and Rosengren and Webster and Phalen, there are many ways to approach the concept of audiences (2014, p. 4 ff.). McQuail adapts the concept of audiences as a more general one and emphasises that the use of the term "audience" leaves – beyond the fact that it is understood by media practitioners as well as scientists–ample space for "differences of meaning and theoretical disputes" (McQuail, 2000, p. 360).

"The term audience, which was and to some extent still is satisfactory for mass media research, fits poorly within the domain of new media, in a number of important ways, audiences are becoming 'users'. […] the term users […] better covers this variety of modes of engagement" (Lievrouw and Livingstone, cited by Carpentier, Schrøder and Hallett, 2014, p.5) but audiences are still sold to advertisers; they are a revenue source for television and social media.

All research approaches have in common the fact that it helps to understand a theoretically constructed audience. According to McQuail (2000, p. 365) audience research goals can be classified as follows:

- Search for a currency for the measuring of reach and value of advertising
- Forming of audience choice behaviour
- Finding audience market opportunities
- Testing of television programs (products)
- Improve communication effectiveness
- Control if the mandate to serve an audience is met
- Evaluating media performance

Table 2.7. McQuail's three audience research traditions (McQuail, 2000, p.368).

	Structural	Behavioral	Cultural
Main aims	Describe composition; enumerate; relate to society	Explain and predict choices, reactions, effects	Understand meaning of content received and of use in context
Main data	Social-demographic, media and time use	Motives; acts of choice; reactions	Perceptions of meaning in social and cultural context
Main methods	Survey and statistical analysis	Survey; experiment; mental measurement	Ethnographic; qualitative

The research of television audiences is as old as television itself and the media industry has always been searching for the ideal methodology and metrics to evaluate audience behaviour. The industry and sciences use many methodologies to better understand television audiences and for decades audience research has methodologically and technologically been in a process of development and improvement to produce as exact data as possible about television audiences. A currency to evaluate the daily broadcasting is needed as an independent source to evaluate the performance of broadcasting, and quantitative and qualitative data about television audiences to understand who likes or dislikes program.

The critical view of scientists charges the television industry, as the complex viewing practice of television audiences is routinely transformed into a "piece of commercial information called 'ratings'" (Ang citied in McQuail, 2000, p. 364). However, the need of television industries for "structural audience research" (table 2.7) to obtain reliable estimates about the size and reach of television audiences is essential; in addition to size it was important to get information about the socio-demographic composition of the audience (McQuail, 2000, p. 368). Perhaps best-known approach in the industry to researching television audiences is the quantitative rating analysis which is performed by companies such as The Nielsen Company (Nielsen), comparable to Arbitron in the radio industry (Buzzard, 2012). Nielsen uses the so-called people-meter-method as a tool to measure the viewing habits of television audiences which will be described in detail in chapter 3. "The people meter in its various forms was the latest evolution of a body of audience measurement methods and tools that came to define the twentieth-century TV and radio marketplace" (Buzzard, 2012, position 99). Measuring the behaviour of television audiences by data collection, processing and delivery based on people-meter-methodology (chapter 3) is applied in most of the television markets all over the world. This method submits on a daily and minute by minute basis relevant key metrics, so-called television ratings and television (market) shares to the players and analysts of the television industry.

As described in this chapter, technology develops fast and audiences are getting more and more fragmented due to a diversity of channels, platforms and other media products such as time-shifted viewing. Audiences have more options of media usage than before and develop new habits.

Although most of the statements in this chapter address internet transformations, the "traditional media" and their audiences are under change (Patriarche, et al. cited Evans, Krotz and Hepp, Lundby and Rudin, 2014).

To avoid the misinterpretation of television ratings it could be asked if verbal statements such as social media comments of the television viewer could increase understanding of the audiences. As argued by de Bens, although there is a high level of trust in television ratings, there is a need to gather information not only on the amount of television consumption which is measured but also on the qualitative aspects of television consumption to understand how audiences consume and understand television. (2005, p. 82).

The aim of commercial and public television broadcasters is to understand the behaviour of television audiences, and to measure if they fulfil their mission to serve and reach the audience. Television audience measurement submits key data of the television markets such as the market shares of television channels in the respective market, evaluation of the strongest television programs in the market, and demographic information about Croatian television audiences. In general, the Croatian television market shows changes which could be interpreted as motivation to revisit traditional audience research procedures and it seems that although traditional methods submitted useful information during recent decades, that it is not enough for media practioners who are consistently interested in more detailed information about television audiences (Vincente-Mariono, 2014).

The analysis of Nielsen data shows that competition between HTV1, Nova and RTL is high and reflected by their similar market shares (figure2.8-2.10). Croatian television viewers are exposed to programing of very high quality as public and commercial television is strong and invests significantly into television programs. The domestic public television channel HTV1 was the television channel which was most watched between 1990 and 2010 (Car and Andrijašević, 2012, p. 24). Although new digital channels have been introduced, the television market is today still dominated by three groups: HTV (public television), Nova TV and RTL Televizija (both commercial TV). All groups need the results of television audience research to analyse their market positions and the success of television programs. The simplest way to compare television stations or television program audiences is to compare them in terms of market shares (Webster, Phalen and Lichty, 2007). Figures 2.9, 2.10 and 2.11 reflect the change of market shares of main television channels in Croatia during the period of 2011 to 2013.

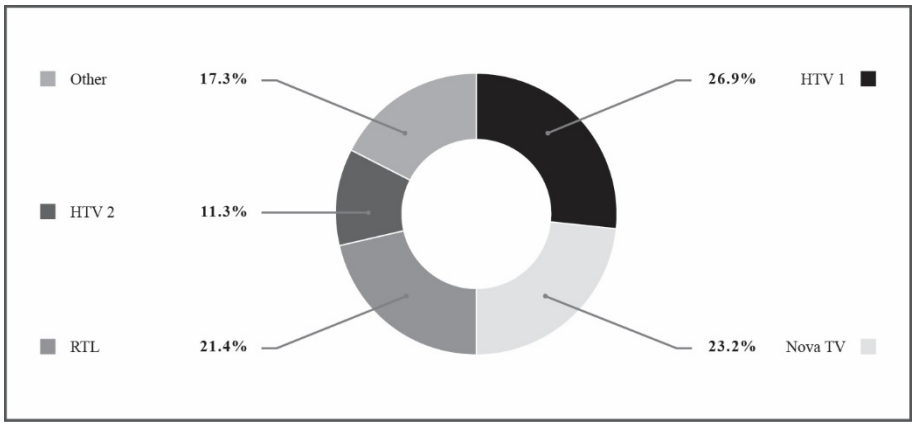


Figure 2.9. Audience shares of main channels in 2011, individuals aged 4+ (RTL Group, 2012).

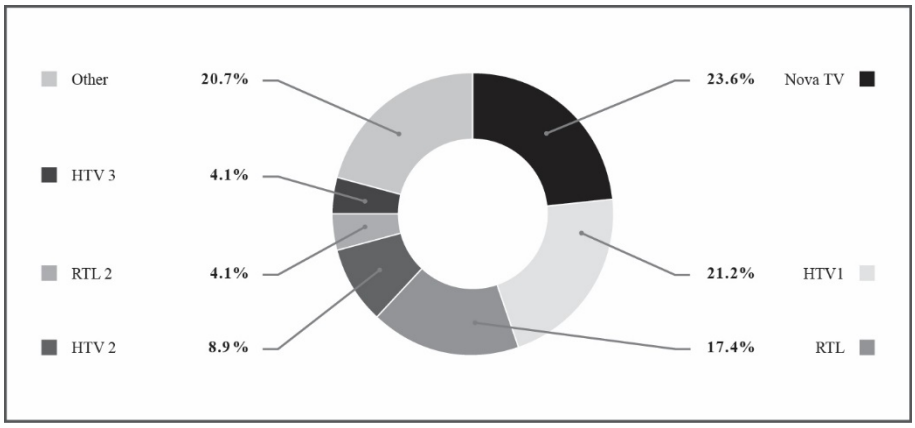


Figure 2.10. Audience shares of main channels in 2012 (RTL Group, 2013).

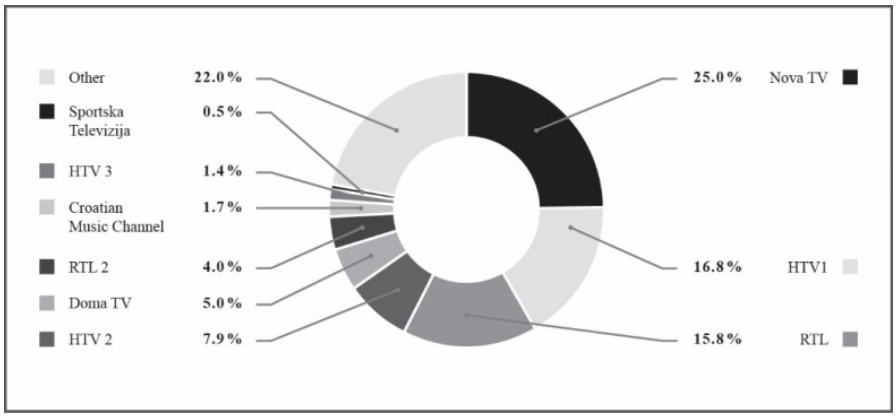


Figure 2.11. Audience shares of main channels in 2013 (RTL Group,2014).

Comparing the numbers it is noticeable that the percentage of *other* channels is significantly growing. The aforementioned fragmentation of the television market is indicated by the increase of the market share of 'other channels' from 17.3% in 2011 to 22.0% in 2013 which means a loss of approximatively 5% for the market shares of the main channels. In 2008, the market share of other channels was 9.2% (RTL Group, 2008), therefore it doubled within four years. Within fragmentation of television markets, the television ratings of main channels and programs decline. In combination with time-shifted viewing, same-day television ratings may not be accurate enough in the future (Jennes and Pierson, 2012).

Historically, the starting point of television ratings research was different to the situation today as there was no free accessible data about the audiences available. There is a

fundamental transformation of audience habits towards media convergence and a "participatory culture" (Jenkins, 2006). With growing social media, the access to a big data open source is apparent, as audiences are everywhere, both as recipients and increasingly as senders of information (2014).

In competition for the attention of the audience, it has always been crucial to know the audience's opinion of individual channels and shows. In the past, the television industry could only indirectly deduce this from the television ratings and a large viewership reflected in high television ratings was equated to the measured show's attractiveness in traditional audience research.

The analysis of social media data could supplement traditional audience research and gain significance within the television industry as social media analysis offers tools to mine user opinions. The present research work approaches the topic of social media analysis in the framework of television research, as this could produce a more precise – albeit not yet representative – reflection of public opinion.

In general, the characterisation of Twitter as a back channel for television is frequently discovered in literature (Bredl, et al., 2014; Highfield, Harrington and Bruns, 2013a; Harrington, Highfield, and Bruns 2013b; Bruns and Stieglitz, 2012), referred to as the quick possibility of expressing a social response. This evidently provides direct data on the opinion of the audience to television research and the question of whether and how this data is evaluable remains a question to be answered. In theory, Twitter data is available in real-time, however, the question remains whether or not it can even be analysed in real-time due to the enormous volume of data.

Jensen (2014) has sought to highlight the changing conditions under which audience research operates in the digital environment and indicates that it would be helpful to draw attention on the diverse communication practices in which media and audiences engage. As media communicate across technological platforms, research can track and rely on growing masses of data, some of which can be found and interpreted. "We need all the methodologies we can devise in order to measure and interpret audiences" (p. 237).

3 CHALLENGES IN TELEVISION AUDIENCE RESEARCH

Research in television helps with understanding audiences and how public and commercial broadcasters' best serve their audience. Regardless of whether a television broadcaster represents its license fee payers or the advertising industry, the views of the audience needs to be understood. The routes and development of television audience measurement has been described exhaustively in literature (e.g.: Ang, 1991; Webster, Phalen and Lichty, 2007; Bjur, 2009; Buzzard, 2012).

The need for reliable and accessible television audience data to analyse the behaviour and habits of audiences has always been great. Not only do commercial television stations target large audiences, public service broadcasting must also be able to provide "a wide range of programming […] in order to address society as a whole; in this context it is legitimate for public service broadcasting to seek to reach wide audiences" (European Commission, 2001). Patterns of content preference are continuously recorded and analysed according to demographics, lifestyles, age groups, education levels, gender, etc. in order to deliver the right content to the right audience. Following this, most working theories consider what content will appeal to which groupings within the audience and in which way television programs and channels could be organised and scheduled to maximise the size of the desired segment of the audience (ESOMAR, Ettema & Whitney and Kent, citied in Bjur, 2009).

Historically, television audience research has developed along two traditions: the quantitative tradition, which is based on standardised methods and surveys, in particular audience ratings analysis, and qualitative traditions which rely on qualitative analysis such as focus groups (Patriarche, et al., 2014).

"Ratings analysis is the analysis of the audience size and composition data produced by audience measurement firms for use in both the commercial and non-commercial media sectors" (Napoli, 2011, p. 286). Television audience measurement submits key television market data, such as the market shares of television channels in the respective market, evaluation of the strongest television programs in the market, and demographic information about Croatian television audiences. It submits information and currency to the industry and is used in making decisions about the launch of new television channels, about the production or acquisition of television programs such as feature films, television series, lifestyle programs,

news, sports, and documentaries, and decisions about the everyday business of buying and selling advertising.

Television channels, program makers and advertisers have an interest in learning about people who engage with and talk about the program. Recent reviews have documented the theoretical discourse about the role of television audiences, the structure of television viewers and the measurement of television audiences (Ang, 1991; Fiske and Hartley, 2003; Morley 2010; Napoli, 2011; Silverstone 2003; Buzzard, 2002; Reinold, 1994; Bjur, 2009).

Television ratings and the idea of audiences are used by all market participants to legitimise their activity, yet the idea of an audience is a myth which can be interpreted and defined more easily than a relationship with a real actor (de Bens et al., 2005).

Since the 1980s and 1990s measurements by the people meter system is the dominating and reliable measurement practice in the industry (Ang, 1991; McQuail, 1997; Bjur, 2009). The tracking of viewer behaviour is used to draw conclusions about the popularity of television channels and programs. People Meter is the official standard in most national television markets (Webster et al., 2007). “That People Meter has become the ideal rating service does not mean that it is a waterproof television audience estimation methodology” (Bjur, 2009).

These numbers are seen as the main currency of the industry and are “of such overwhelming importance to the media, that some textbook writers simply draw a distinction between ratings and non-ratings research” (Wimmer and Dominick, 2011, P. 352). Ratings data are used by policy makers to assess and evaluate media markets and by academics to better understand how audiences consume media (Napoli, 2011).

The advantages of people-meter methodology as a structured approach to television audience research are demonstrated when the fundamental requirements and expectations of the industry are met.

Television ratings measurements made with people meters are still the most valid and trusted common currency of the television industry and the most advanced system of measuring the viewing behaviour of television audiences (Bjur, 2009). The accuracy of data may vary depending on providers and system design, but the system as a whole has been a reliable tool for the entire industry (Bourdon and Méadel, 2011).

The presence of panel members is tracked and extrapolated to the population of television households. The panel members use a remote control to communicate with the system

and to register when they tuned into a certain program and channel. The result of the people-meter measurement is finding out "who watched which program and channel and when". Furthermore, it provides information about the demographic character of the viewer so that the results can be analysed from different angles, such as the popularity of certain programs among different audience groups (gender, age, income, and region). The sampling, statistic collection and generalisation of data and the broad interest in the audience implies an analytical transfer from individual households to a general audience.

High numbers of TV ratings and market shares are considered to represent successful and popular programs which television audiences like and enjoy viewing.

3.1 Television audience analysis in context of history

The television market in the United States has played a pioneering role in the development of audience research, and with 114 million television households (Statista, 2014), it is the largest and the oldest in the world. The development of television ratings measurement is related to a prosperous advertising market centred on radio and television at the beginning of the nineteenth century in the United States. Public television started in 1954 with national educational television and was substituted in 1970 by the public broadcasting system, but it never had the important role comparable to that of public broadcasting in many European countries. The initiators of television audience research were the television broadcasters – not the advertisement industry – and it has always been the main interest of broadcasters to get information about audience behaviour, size, and structure (Wimmer and Dominick, 2011 p. 351).

The commercial networks NBC, ABC and CBS in the United States sought audience research at the beginning of the 1920s, as radio stations and the advertising market became more powerful and organised, and the demand for the establishment of audience measurement remained. In the beginning of television, information about media audiences consisted of "subjective impressions such as anecdotes, postcards mailed in by audiences, and other schemes conceived by the advertisers" (Buzzard 2012, position 112). The economic operation of television, radio, and the internet requires information that buyers and sellers of advertising need to agree to use. The aim of commercial mass media as sellers of advertising is to maximise their profits by reaching maximal audience sizes in special demographic groups.

The research departments of newspapers started collecting data in 1914 about the purchasing behaviour of households and circulation figures for advertisers. When launched radio services did not know who was listening to their programs. Needing a valuable business model, they were forced to know more about their anonymous radio audience. The media and the advertising industry had to address the problem that an independent institution was needed to provide professional information about audiences and that a common currency had to be found.

One of the pioneers in the field of television audience research was Arthur Nielsen, an American market analyst who founded the A.C. Nielsen Company (Nielsen) in Chicago in 1923. Before television became popular, Nielsen used audiometer radio ratings, to measure which radio station had been listened to by a household during the day. In generating data about audiences, Nielsen took part in a new growing sector of the economy: the development of an information industry where information about audiences becomes the central part of the economic activities of mass media. Nielsen's approach to audience measurement was based on his belief in sampling and that by collecting data from samples it could systematically and statistically provide a picture of a society.

In 1952 Nielsen entered the television market and developed a television audience measurement system to measure and to statistically estimate the audience size of television programs based on a sampling of homes across U.S. This was the beginning of television audience measurement and is today commonly referred to as television-ratings measurement. Initially, Nielsen sent out paper diaries to record viewing, and then later introduced the boxes in selected viewers' homes which automatically measured the programs that viewers watched. Nielsen used the electronic metering technology to gather random samples of a defined number of households in the individual country which reflected demographic distribution nationwide.

Over the decades other players came into the audience measurement research market to provide international television audience ratings. Due to mergers and joint ventures the research market of today is highly concentrated. The Nielsen Company went public in 2011 and is listed on the New York Stock Exchange. Nielsen is present in Asia, Europe, the U.S.A, Africa, and South America. In 2005 AGB Group and Nielsen Media Research merged to become AGB Nielsen Media Research. Most of the audience data in Europe is provided by AGB Nielsen Media Research, Taylor Nelson Sofres, and GFK Group.

"By the late 1980s the people meter had undeniably become the industry's token of what has been called a ratings 'revolution'" (Ang 1991, p.63). The traditional audience measurement with people-meter systems – although often criticised – has proven to be the reliable method for the television industry over the last three decades (Bjur, 2009) and has been by far the most relevant method. The survey which is conducted to recruit television households into the panel and the case study which is defined to analyse the behaviour of television audiences focus on the following research questions: "Who watches what, at what time and for how long" (Bjur 2009, p. 85). It involves the decision of the audience of whether to watch or not, and whether or not to select a specific show on a selected channel (Tavakoli and Cave, 1996) based on whether or not the television set was on or off. Over the last 15 years, the methodology of providing data about television audiences using the people-meter system became a standard for most of the countries in the world, and the system produces television ratings and television shares in all television markets.

Today, audience measurement is conducted electronically with a dual metering system, so-called 'people-meters', which are installed in a statistically representative sample of homes on every television set in the panel home, to represent the national universe of television households. Depending on the provider of television research analysis and individual countries, the data production includes different specifications and has different accuracy. With the advent of digital television and the use of new technology, such as digital video recorders, set-top boxes, tablets, and mobile phones, media players needed to find new ways to submit content and new methods to track media consumption. Nielsen invested major efforts in its data delivery to submit more detailed data and had to expand its samples sizes to deal with fragmentation and the differences in the way audiences watched television (Buzzard, 2012, position 197 and 1643).

Audience measurement with people meter technology in Croatia is performed by AGB Nielsen Research. The electronic measurement panel in Croatia was launched in 2002 and AGB Nielsen Media Research adapted people meter technology to the Croatian market. In the following section, the practical organisation of audience measurement provided by the audience analysis business, and outlined by Webster et al. (2000), is presented.

3.2 Description of television ratings methodology

Every country or television market needs valid and reliable data on television audiences. Essentially, the basic idea of the people meter as an electronic monitoring device that can record individual viewing is simple (Ang 1991, p. 63). The people-meters detect and register behaviour of the sampled households concerning television viewing by using remote control sets which are to be pressed when the viewer is in front of the television. The sample of households is representative based on demographic features of the Croatian population and the characteristics and penetration of the television equipment of the households. The selection of television households is random, which means that there is a systematic random selection of houses, locations and streets, so that every household has the same chance of being chosen (Stiller, 2005, p. 6). The people-meter registers the on/off timing of the television set, the selected television equipment for viewing (television set, VCR, satellite or cable receiver) and information about which channel and program was selected. The information remote control of the measurement system requires input action from the panel member by using the remote control (Bjur, 2009). The remote control registers:

- which member (age and sex) in the household is viewing,
- the arrival and leaving time of the viewer,
- information about departure of the panel household e.g. for a holiday.

The audience ratings research is interested in target groups, size of the audience, and main demographic characteristics. The television ratings have an influence on the program acquisition and production of the television station, the scheduling of television programs, and the set of advertising prices (rates). A successful television program or television station is thus defined as one which has high television ratings and/or market shares.

3.2.1 Description of people-meter methodology

To guarantee reliable, independent and transparent data production, people-meter methodology includes a data system that collects and analyses data of television programs and advertisements to measure the television audience. It includes program viewing, interruptions,

and television commercials. For the data collection and reporting, an ideal procedure and configuration of people-meter-process has to be established figure 3.1, and regardless of the type of organisation or data supplier submitting the analysis, the formal procedures have to be followed and all aspects of research methods must be transparent.

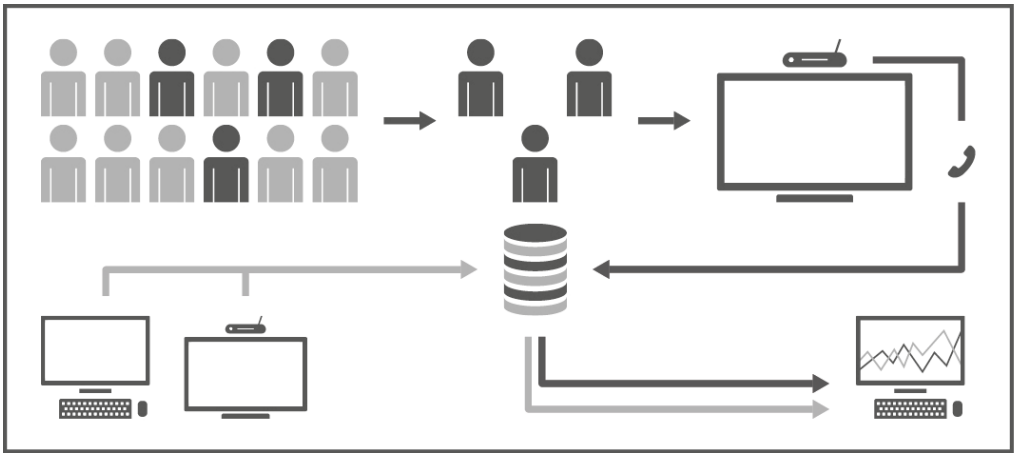


Figure 3.1. People meter configuration (Bjur, 2009; Webster, 2007).

Such transparency ensures that all market participants can trust the data. The methodology for audience research based on people-meter includes in four general steps:

Step 1: Establishment of a survey to select panel households

The approach of this methodology is based on a sample survey. An extensive process precedes measurement of television data as the sampling begins with the selection of households and the quality of the sample "has a tremendous impact" on the accuracy of the collected data (Webster, Phalen and Lichty, 2007, p. 112). The objective of the survey is to collect the demographic information of potential panel households and to select the sample of households (panel members). The survey includes households based on 2011 census data and to ensure valid representativeness of the population.

The panel has to represent the Croatian television population in respect to demographic and geographic characteristics and technical equipment. The presumed representativeness of these samples allows for projections of the Croatian population as a whole. Generating accurate samples is challenging because the survey results are influenced by various issues: for example, many potential participants do not wish to participate. The survey is conducted via face-to-face

interviews. According to Nielsen, the first establishment survey in Croatia was conducted using a sample poll of 6,600 families (10 to 1 in terms of panel size, N = 660 families). The composition of the panel is (partially) renewed on a yearly basis to ensure that the sample represents the Croatian television audience. As well as the quality of the sample, the size of the panels is the second key factor because the sample represents the Croatian television population. The sample provides information about the character and the structure of the national television population which could not be researched thoroughly, due to financial reasons, or due to time constraints or ethics issues (Sachs, 2002).

Step 2: Panel of television households in Croatia

Given requirements of different television market, a panel size can range from 5,000 households in major markets such as the USA (115 million television households) and Germany (38 million television households), to approximately 500 panel members in smaller markets. Currently the panel size in Croatia is composed of 810 households (it increased from 760 in 2009). This sample represents a population of 1.4 million television households in Croatia with an average of three household members. In Croatia every sample household represents 1,777 households in the country. There are 1,200 people meters in 810 households and 2,500 individuals aged 4 or above living in a household with at least one television set participate. As can be seen in table 4.1, the sample size of smaller countries such as Croatia and its neighbours is relatively high compared to other television markets in neighbouring countries. The biggest panel in Europe is in Germany with a sample size of 6,926 panel households.

Table 3.1. Panel Size (Nielsen, 2013; wikipedia.org).

	Population (in 000)	TV households (in 000)	Panel Size (households)	Number of households represented by one people meter
Hungary	9,733	3,953	1,040	3,801
Croatia	4,437	1,440	810	1,778
Serbia	7,187	2,465	880	2,801
Slovenia	2,059	707	450	1,571
Bosnia & Herzegovina	3,834	1,166	550	2,120
Germany	**81,844**	**39,065**	**5,640**	**6,926**

Based on geographical representativeness Nielsen (2013) divided Croatian territory (figure 3.2) into six regions: Zagreb (1), Northern Croatia (2), Slavonia (3), Lika and Banovina (4), Croatian Littoral and Istria (5), and Dalmatia (6).

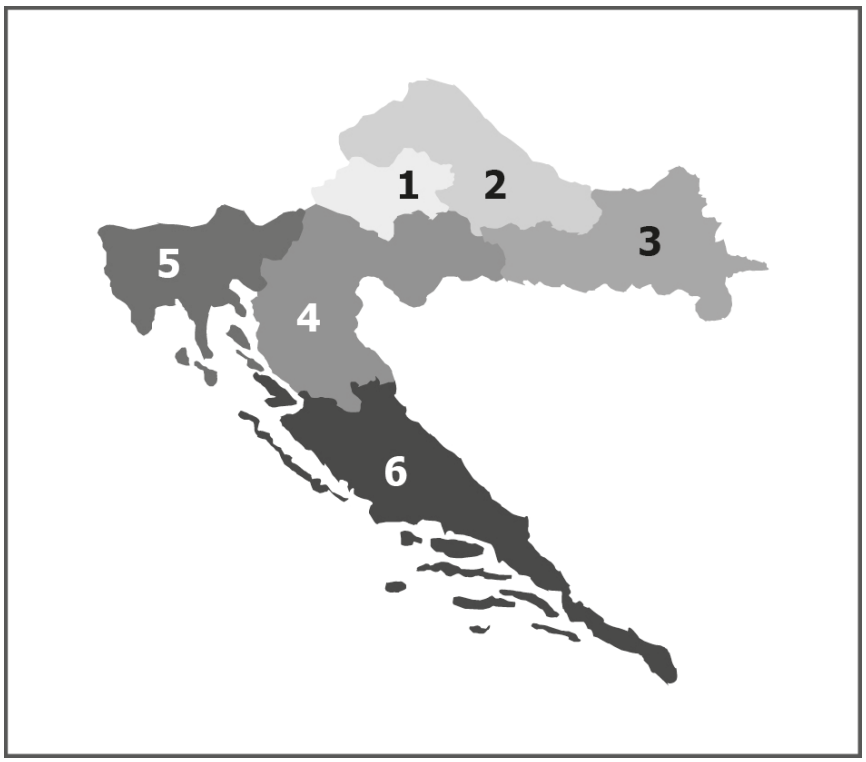


Figure 3.2. Map of Croatia, regions 1-6.

There are 810 television panel households distributed all over Croatia and the selected households are characterised by household size and composition, region or city size, technical equipment, as well as additional information about the household such as its members' level of education, social class, and income.

Step 3: People-meter-device

Every selected panel household is equipped with a people-meter, which is an audience measurement tool connected to the digital television receiver of the household and which can collect electronic data. The participants must log in and out each time they watch television so that the people-meter registers the exact time of viewing and who is viewing. The viewer uses a special remote control, so that information about the channel that is watched is tracked automatically. Every broadcasted program and commercial break are entered into the system

including the exact start and end time of television programs so that it can accurately calculate the size and composition of the audience of an individual program.

Step 4: Data transfer, collection, processing and analysis of audience data

The people-meters register the television status information at an interval of one minute, for 24hours a day, every day. Via modem telephone lines or the internet, the protocol of data is retrieved every night to poll and validate the data to a central computer. The input is processed by a computer application which generates the television ratings data.

3.2.2 Sociodemographic data, television ratings and market shares

Measuring and researching the behaviour of television audiences is the working foundation of the whole television industry. The simplest way to compare station or television program audiences is to compare them in terms of market shares (Webster, Phalen and Lichty, 2007). Furthermore "a potentially overwhelming amount of demographic information" (Webster, Phalen and Lichty, 2007, p. 163) is submitted by using the people-meter methodology. To understand the value of traditional audience research it is import to understand what kind of information it generates. In particular, the quality of demographic information which is included with key metrics shall be investigated as the combination of the characteristics of the panel member with television ratings is interesting to the researcher. It will be seen that the characteristics of the panel members are (partly) qualitative statements about the television household and that this allows a mixture of quantitative and qualitative (Berger, 2015) conclusions about the television viewers to be drawn. The key metrics determined and reported based on the collected data are the ratios of television ratings and television (market) shares (Webster, Phalen and Lichty, 2007).

Television ratings for every program and television station are available on a minute by minute basis. A television rating is the percentage of all households who tuned into a selected television program and stayed on the channel for more than one minute. This means that the percentage refers to a number of households as a percentage of the entire television population. The market or broadcasting share is defined as the percentage of television households where the television set is on during one particular period, meaning that the number refers only to the percentage of the entire television population that was watching a particular television program

at that given moment. Television ratings and market shares (figure 3.3 and figure 3.4) are explored and reported in different ways, depending on criteria such as definition of target groups, and time zones. Expressed as formulas, television ratings and market shares are calculated as a ratio (Webster, 2007):

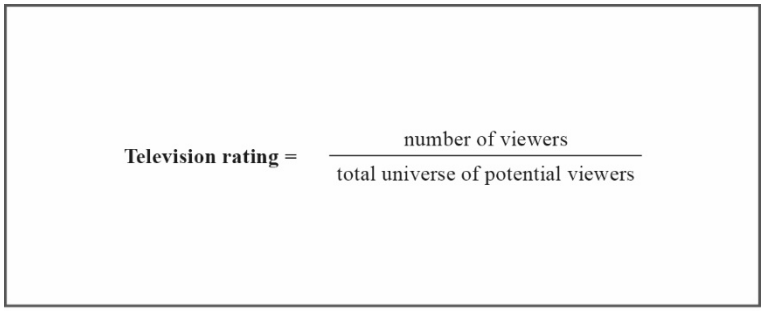


Figure 3.3. Television ratings.

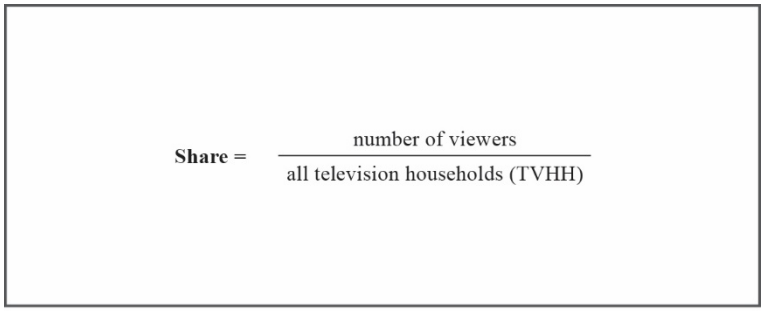


Figure 3.4. Television market shares.

There are different methods of representing the results of people-meter measurement. Market shares and television ratings can be calculated per television channel or for single television programs. Given that the people meter provides continuous data over a long period of time, it is possible to accumulate information about the viewing habits of television viewers.

The peak time of television which means the time when most of people in Croatia watch television is 9:15 pm. Most people watch on Sunday evenings (figure 3.5).

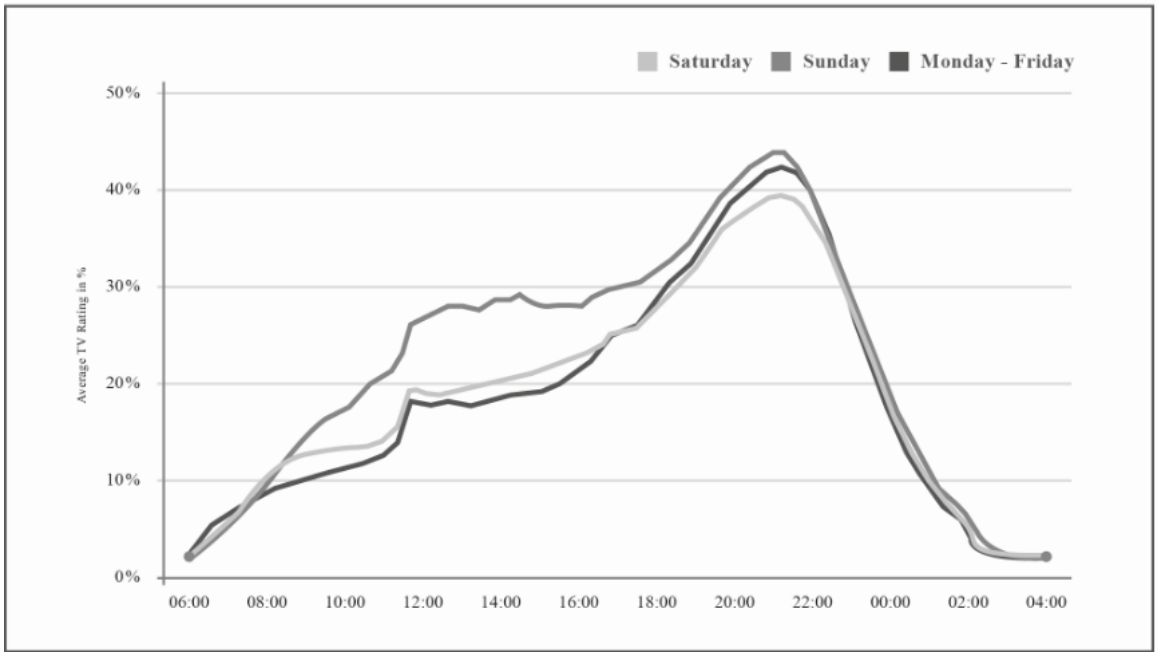


Figure 3.5. Viewing patterns in Croatia 2013. (RTL Group, 2014).

In respect to the most popular television programs in Croatia, table3.2 shows the ratings and market shares for 2013. In that year football and handball dominated the Croatian television market. Table 3.2 shows the ranking of television programs which is important to the television channel operators as it influences the share of advertising for the individual channel and forms the basis for making program decisions.

Table 3.2. Market shares and television ratings of the most popular television programs in Croatia 2013 (AGB Nielsen Research, 2013).

#	Title of Program	Channel	AMR %	SHR %
1	Nogomet - Kvalifikacije ZA SP (DOM) - Prijenos	HTV2	29.49	65.78
2	Nogomet - Kvalifikacije ZA SP - Prijenos	HTV2	29.06	68.41
3	Nogomet - Kvalifikacije ZA SP (DOM) - Prijenos	HTV2	28.28	54.42
4	Rukomet - SP (M) - Spanjolska - Prijenos	HTV2	25.79	54.68
5	Nogomet - Kvalifikacije ZA SP - Emisija	HTV2	23.70	54.99
6	Lud, Zbunjen, Normalan - Humoristicna Serija	NovaTV	23.10	49.72
7	Nogomet - Kvalifikacije ZA SP - Prijenos	HTV2	22.23	48.48
8	Nogomet - Kvalifikacije ZA SP (DOM) - Prijenos	HTV2	21.92	50.93
9	Rukomet - SP - Emisija	HTV2	21.73	58.11
10	Dnevnik Nova TV	NovaTV	21.52	46.76
11	Legend II Invasion - Snimka	NovaTV	21.49	52.06
12	Nogomet - Kvalifikacije ZA SP - Prijenos	HTV2	21.38	44.51
13	Sila	NovaTV	21.29	43.44
14	K - 1 World Grand Prix 2013 Final - Prijenos	NovaTV	21.13	56.68
15	Stella - Hrvatska Glazbena Dramska Serija	NovaTV	20.11	40.69
16	Rukomet - SP (M) - Spanjolska - Prijenos	HTV2	19.24	40.96
17	Dnevnik U Podne	HTV1	19.17	57.45
18	Legend II Invasion - Snimka	NovaTV	19.04	43.39
19	Plodovi Zemlje - Emisija O Pojoprivredi	HTV1	18.75	52.29
20	Legend II Invasion - Snimka	NovaTV	18.32	42.92
21	Djevojka Imena Feriha - Turska Dramska Serija	NovaTV	18.28	39.79
22	Boks - Teska Kategorija (EUR) - Prijenos	NovaTV	18.26	44.92
23	Nogomet - Liga Prvaka - Prijenos	HTV2	18.22	45.24
24	Legend II Invasion - Snimka	NovaTV	17.70	39.95
25	Rukomet - SP (M) - Spanjolska Prijenos	HTV2	17.08	34.55
26	Nogomet - Kvalifikacije ZA SP - Prijenos	HTV2	16.94	39.74
27	Nogomet - Kvalifikacije ZA SP (DOM) - Prijenos	HTV2	16.66	44.23
28	Rukomet - SP (M) - Spanjolska Prijeno	HTV2	16.50	35.29
29	Sulejman Velicanstveni - Turska Dramska Serija	RTL	16.14	32.77
30	Vijesti Nova TV	NovaTV	15.92	34.46
31	Zora Dubrovacka - Hrvatska Dramska Serija	NovaTV	15.84	35.75
32	Larin Izbor - Hrvatska Dramska Serija	NovaTV	15.72	35.94
33	Ljubav Je Na Selu - Dokumentarna Sapunica	RTL	15.22	32.80
34	Nedjeljom U Dva - Politicka Emisija	HTV1	15.12	41.27
35	Dva Smo Svijeta Razlicita - Humoristicna Serija	NovaTV	14.95	36.86
36	RTL Danas	RTL	14.84	38.11
37	Ples Sa Zvijezdama - Zabavna Emisija	HTV1	14.75	34.26
38	I Godina Nova 2014	NovaTV	14.73	36.64
39	Sam U Kuci 3 - Americka Komedija	RTL	14.16	30.82
40	Rukomet - SP (M) - Spanjolska - Prijenos	HTV2	14.06	29.04
41	In Magazin	NovaTV	13.97	38.53
42	Provjereno	NovaTV	13.94	34.43
43	Kako Vrijeme Prolazi - Turska Dramska Serija	NovaTV	13.92	36.02
44	Nogomet - Hrvatski Super Cup - Projenos	NovaTV	13.90	41.05
45	Sam U Kuci 2 - Americka Komedija	RTL	13.87	33.01
46	RTL Vijesti	RTL	13.87	35.02
47	Rukomet - SP (M) - Spanjolska - Prijenos	HTV2	13.64	28.65

Television audience measurement submits key data of the television markets such as the market shares of television channels in the respective market, evaluation of the strongest television programs on the market, and demographic information about Croatian television audiences. The research of Nielsen includes qualitative sociodemographic information as a result of the establishment survey which can be combined with key metric numbers such as market shares and television ratings.

In general, television consumption depends on several circumstances such as the time of day, the season, and other activities, therefore the market share is often considered to be more meaningful because it provides information about the television households which actually have their television sets tuned on. Television ratings are coupled with demographic information and characteristics about the family such as income, education, and lifestyle to reveal if the viewing habits are related to certain characteristics. With people meter data it is therefore possible to reveal, for example, the percentage of women or men aged between 18-34 who watched the first half of a football match, or, during an advertising commercial break, which demographic target group has seen an advertising campaign (compare Ang 1991, p. 64).

There are 810 television panel households distributed all over Croatia and the selected households are characterised according to household size and composition. The characteristics of the participating panel households can be combined with the key metric data (ratings/market shares). The population of television households that is being measured and reported is defined by a selection of demographic, geographic, housing, equipment and other criteria. The combination of different data could enable the researcher to draw qualitative conclusions in respect to the audience.

In Croatia, 202 of the television panel households are based in Zagreb, which means that 25% of the panel measurement in Croatia takes place in the capital city. The people meter panels are established from large surveys defining the composition of the television audience universe (Bjur, 2009, Miketek, 2014). The panel households are clustered according to defined criteria (Miketek, 2014):

The panel households of the Croatian television panels are distributed between areas with more than and areas with less than 5,000 inhabitants. It could be interesting to research which television programmes or television channels are more popular in cities with various population sizes. The combination of this characteristic of audience data with television ratings and market shares may allow for conclusions regarding whether a particular program,

advertisement or television channel is under certain characteristic conditions more liked by television viewers.

The survey collects audience data to complement data from the national census (2.3). As the cultural and ethnological differences between the regions are known, it is possible to analyse which region (e.g. in provincial or urban environments, figure 3.2) particular television programs or television channels are more liked and accepted than in other regions.

The survey furthermore investigates the technical equipment in television households and gives information if a household is equipped with one, or multiple television sets at home extra digital television satellite and cable channels, so that the question can be discussed if the household equipment does influence the social context of television consumption.

The households are asked within the survey if they have children or not and if they live in the household and if they are dominated by more male or females members.

The survey asks questions about the age of television household members (table 3.3). This is considered to be the main information for advertising clients as age is a significant indicator for consumption behaviour.

Table 3.3. Age clusters of the Croatian television audience.

Gender	Total	Age	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90+
All	4.284.889		417.026	479.579	550.724	579.373	594.494	632.320	474.742	387.927	155.745	12.959
Male	2.066.335		214.092	245.614	280.871	293.982	296.049	311.731	217.215	155.368	48.414	3.062
Female	2.218.554		202.934	234.028	269.853	285.391	298.445	320.589	257.527	232.559	107.331	9.897

The term life style is used informally to point out differences in the way people live and is assumed to be an important variable for studying consumer behaviour (Williams, 1972). The purpose of researching lifestyle in this context is to analyse factors that contribute to television viewing habits.

Table 3.4. Lifestyle of panel households.

1.	Unknown
2.	Hard life
3.	Routine based
4.	Familiy oriented
5.	Loaded with obligations
6.	Open minded

The survey asks television households questions regarding their income and their employment and education status. The answers are categorised into low, middle, upper, and higher income households. Individuals within a social class learn specific values primarily through their associations with family, friends, and neighbours, and in school and at work. In English-speaking countries a blue collar household is working class, where people typically perform manual or low-skilled labour, in contrast to a white collar household where its members typically have high-skilled jobs in an office environment.

Table 3.5. Classification of social classes of panel members (Nielsen, 2013; Žorž Miketek, 2014).

Classification of Social Classes
Rural lower class (retirement age households)
Urban lower class (retirement age households)
Rural middle class (family type households)
Urban lower-middle class (blue collar households)
Urban higher-middle class (white collar households)

It has to be emphasised that the combination of qualitative and quantitative data is considered in literature as a way to validate and deepen the knowledge about television audiences (Webster, 2007).

3.3 Limits of traditional methodology

Although the people-meter method is considered to be the main approach for researching television audience behaviour and an applied currency for the industry in most television nations in the world, the practice of television ratings measurement is often evaluated and discussed by the profession and within the sciences. Researchers accept the "ratings-industry" in general (Bourdon & Méadel, 2011) and "one could write that ratings give a decent approximation of television viewing, as defined at a certain time, in a certain context, under certain requirements, by and for certain actors", (p. 9).

Morley argues that "observing behaviour [of a TV viewer] always leaves open the question of interpretation" (2010, p. 172). Should the researcher wish to better understand television viewer behavior, but does not want to simply reduce the observation to the question of whether or not the television set is on or off, more interpretation and data are necessary. "What is needed is not simply improved techniques of audience measurement (pace the 'passive-peoples' debate) but improved methods of audience research […], so that we cannot only measure what different types of audience do, but also understand how and why they do as they do" (Morley, 2010, p. 167; Wober, 1981, cited in Morley).

Furthermore it is generally presumed that "the audience is measurable and "accounted" for in the literal sense of being expressed as numbers or proportions, whether actual or estimated" (McQuail, 1997, p. 57).

The people meter system methodology reports if the television set was on. The system is designed to track behaviour but cannot submit information regarding whether or not the member of the television-panel household actually followed the show. Furthermore, there is no information about a positive or negative opinion of the television viewer.

The fact that the television household tuned into the program is equated with the assumption that the television household watched the show and was not doing anything else at the same time. The panel viewers must manually tell the device that they do not want to watch television anymore. What also seems to be problematic is the panel's representativeness of the national television audience. The composition of the panel in respect to size and audience is crucial and if the panel is not representative and the rate of panel responses is low, even increasing panel sizes cannot solve the problem (Bjur, 2009). To improve the quality of data, the number of panel households in Croatia has been expanded over recent years. The people

meter data is subject to measurement errors as technical problems may also arise as a result of the reliability of the panel members. How could it be verified that the household member really pressed the button of the panel remote control on time? Regional problems arise, such as long summer breaks in Croatia, and absence of panel households. Measurement is not performed in summer residences during summer time. In Turkey, Nielsen tried to remedy this problem and started the first summer viewing measurement project in 2008, where the people-meters were installed in the summer residences of the households.

Only in some countries portable people meters are used. Besides the methodological challenges of the people-meter, with the increasingly digital and fragmented television environment, audience measurement research is facing new demands. The developing of the internet and the transformation of media habits, have led to reduced television consumption numbers worldwide based on television ratings information. The consumption of television programs has become more mobile and non-linear.

With the introduction of digital set-top boxes new data was available to the researcher providing information via the return-path channel of the box, virtually down to the second, even for the most fragmented television markets. This raised the question about whether census based methods of television audience measurement rather than sample based data of television tuning behaviour could be a better way of gathering data about audiences (Buzzard, 2012). People meter systems are not capable of handling massive fragmentation while set-top boxes could gather data (Buzzard, 2012). Not only was the fragmentation essential, but also the habits of audiences and how people watch television. Television consumption on different devices is more and more typical for the younger generations, who are even starting to have no television set at home. How could the phenomenon of mobile television consumption on second and third devices such as computers, smartphones, and tablets be addressed?

Nielsen plans to expand panels and methodologies and began in 2009 by including internet in the television ratings currency (Buzzard, 2012). They started in the US, where internet viewing and time-shifted television consumption already play a signification role. While initially the people meter methodology could be easily adapted and established in different media systems, the fragmented digital television markets is getting more diversified and national differences demand different television audience measurement methodology, depending on development of the individual television markets.

Although the people-meter method is often critised, at the moment there is no equivalent alternative which would submit data at the same standard regularly to the industry. Lotz (2007) argues that television at the turn of the twenty-first century remains as important as ever but has fundamentally changed as the result of technological innovations, proliferating channels in fragmented markets targeting niche audiences, and new forms of advertising. Therefore, many conventional practices and even basic business models are proving not suitable resulting in a crisis of norms and practices.

Professional audience research is developed and belongs to industry and business, (Bjur, 2009), continuously producing data of audience behaviour for commercial purposes, while the other source is the audience research produced in academia (Weibull and Webster et al., citied in Bjur, 2009). "Knowledge production from both these sources have, since the advent of broadcasting, built our present image of audience behaviour and have provided provisional truths regarding when, how and why individuals engage in media use and media content consumption" (p. 28).

Is it possible to provide a clearer picture of the "present image" of audience behaviour with social media analytics? The following chapter will investigate this question: if social media activity in the context of television can be analysed in a way that could make it an additional source of information to better understand television audiences.

4 RESEARCHING SOCIAL MEDIA ACTIVITIES

Industries can use social media monitoring to track opinions about products and brands, to respond to consumer insights and modify their marketing messages, for brand positioning or the development of their products (Zabin and Jefferies, 2008; Agrawal et al., 2003; Bautin, Vijayarenu and Skiena, 2008; Benamara et al., 2007; Fukuhara, Nakagawa and Nishida, 2007; Godbole, Srinivasaiah and Skiena, 2007; Kale, et al., 2007; Somasundaran et al., 2007; Ounis, Macdonald and Soboroff, 2008). Social media platforms enable the creation and exchange of content and what they all have in common is that they define the rules of how the data is structured, stored and accessible to the public. The amount of data has been dramatically increasing, and analysing large data sets - so-called big data - will become a key basis of competition, and leaders in every sector have to manage the implications of this (Manyika, Chui, Brown, et al., 2011; Mayer-Schönberger and Cukier, 2013).

The rise of big data is a socio-technical phenomenon and there is no doubt that the quantities of data are enormous. "Big Data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets" (Boyd and Crawford, 2012, p. 663). Big data offers to researchers new ways to claim objective research methods and data analysis but one still has to bear in mind that social media as a data source has to be interpreted carefully as "working with Big Data is still subjective, and what it quantifies does not necessarily have a closer claim on objective truth – particularly when considering messages from social media sites" (p. 667). Although numbers of social usage are high, statistical representativeness of data for the (television) audience is generally not given, as not every television household (2.3.1) uses social media (2.2.3).

The consumption of mass media has always been a collective experience, but what is now a feature is the fact that these experiences can be shared digitally among anonymous or partially anonymous audiences. People communicate and share their television experiences and they leave traces and information about their viewing behaviour, "there is an added incentive for us to talk among ourselves about the media we consume. This conversation creates buzz that is increasingly valued by the media industry" (Jenkins, 2006, p. 4). Audiences have become increasingly observable and less abstract; scholars are able to learn more and know more about audiences than ever before. The possibilities of audience research and techniques have changed over the last years, "a new type and scale of data are there to be found: big data or *metadata* that indicate who did what, with which information, together with whom, when, for how long

and in which sequences and networks" (Jensen, 2014, p. 229). With the advent of audiences using social media parallel to television consumption, it can be assumed that the data situation has changed. "There is a new, broad field for television audience research emerging based on the interwoven use of television and Twitter" (Bredl et al., 2014, p. 197). Analysing social media data is still a new field in the social sciences. Social networks and microblogging services are the data houses of our age and can be accessed – depending on the definition of their data policy (Züll and Mikelić Preradović, 2013).

Researchers are confronted with the question of which methods and strategies can be used to understand television audiences and if it would be possible to adapt some aspect of traditional audience research methods to new online data (Bredl, Hünniger and Jensen, 2012).

Besides questions of methodology, there are questions of reliability of available tools in being able to handle the volume and velocity of data and technological challenges for researchers. The crucial question within the framework of this work was how to approach or even how to get access to the global memory as a data warehouse as a basis for the analysis of social media data (social media analytics), because research can only be as good as its data sources are. All users of social media leave digital traces on the internet, and once uploaded, content is difficult to delete (Züll, 2013). Jensen draws attention to the fact that "data can only be found through a good deal of making-programming, extracting, analysing, etc. which suggests that the distinction between found data and made data must be carefully differentiated in each case, it seems clear that the very nature of data calls for reconsideration, in audience studies in general and in methodological literature in particular" (2014, p. 229). The existing question seems to be if the methodological possibilities of the internet to research television audiences can be used. It shall be researched if the comments of television viewers and the online evaluation of television programs can be studied. The focus will be on television audiences as well on television broadcasters as they are the main actors of the media industry and stimulators of online communication. Social networks and microblogging platforms such as Facebook and Twitter and platforms which allow the user to upload media content such as YouTube and Flickr allow the user to participate and communicate related to their media usage.

The main advantage of Twitter is initially the researcher's ease of access to data. Additionally, the characterisation of Twitter as a back channel for television is frequently discovered in literature which refers to the quick possibility of expressing a social response. This evidently provides direct data on the opinion of the audience to television research and the

question of whether and how this data is evaluable begs to be answered. In theory, the Twitter data is available in real-time, however the question remains whether it can even be analysed in real-time due to their large volume of data. A positive aspect is that Twitter provides 'real' data, as no artificial research setting is created here.

"Twitter is a particularly useful source of social media data: using the Twitter API (the application programming interface), which provides structured access to communication data in standardised formats) it is possible, with a little effort and sufficient technical resources, for researchers to gather very large archives of public tweets concerned with a particular topic, theme or event" (Burgess and Bruns, 2012).

Nevertheless, the access to the Twitter data creates vast possibilities for data collection and analysis with relevance in the field of television audience research (Bredl, e al., 2014). In general, user data, (organic or paid) reach, comments, activity on a platform, likes and the analysis of digital traces are the commercial currency of social media platform providers.

The television ratings measurements made with people meters are still the most valid and trusted common currency of the television industry and the most advanced system of measuring the viewing behaviour of television audiences (Bjur, 2009; chapter 3). The accuracy of data may vary depending on providers (Milavsky, 1992) and system design, but the system as a whole has been a reliable tool for the entire industry (Bourdon & Méadel, 2011).

The growth of social media invites many questions as social media have created opportunities to study social data in new ways and massive amounts of television audience data leads to a fundamentally new digital approach in Human Sciences (Manovich, 2012).

Social media analytics in the context of television consumption will be researched, to determine if they can be a new valid source of information. Three hypotheses were deducted when this research was started:

The first hypothesis of this research is that social media plays an increasingly important role for the television audience, giving people the opportunity to share their viewing experiences in various forms and on various social media platforms.

The second hypothesis is that a relationship can be found between television consumption and social media activity.

The third hypothesis is that social media activity can impact on television ratings, and therefore influences the success of a television program.

Social media data can be used as a source for studying how people communicate and interact. Is it possible to take advantage of the new data sources for television audience research? Social media plays an increasingly important part as platforms for public communication and Twitter functions as a back channel for television. The availability of this data raises many questions (Bredl, Hünninger and Jensen, 2014).

The following research tasks will be approached as well as step by step selected aspects with the aim to develop a methodological framework (chapter 6):

- It is impossible to analyse the content of tweets systematically (e.g. text/opinion; links), as finding information about content of tweets could mean getting information about television viewers‘ opinion about a television show.
- To find information about the timeline (time of conversation, before, during or after the show) could mean learning about the general behaviour of television viewers (e.g. do they watch alone or (virtually) with others) and their general viewing habits.
- To find information about the quantity of activity statistical computational approach would help to evaluate the statistical relevance of the results.
- Questions about suitability of television events as a Twitter topic accompany this research.
- To find information about the usage would help in learning which part of the television audience uses Twitter to communicate about television.

The question the researcher faces is not how to obtain audience data but how to use and work with empirical data produced by social media tools which deliver more information about audiences than the television market had ever before. To find answers to the above mentioned questions could give new insights into television audiences and could therefore be a scientific approach to complement traditional television audience research in Croatia with new aspects of social media analysis. Additionally the prevailing usage of social media, particularly Twitter, is still limited to younger people while big parts of the television audience are therefore excluded from this new form of communication. As a result the found data situation (via

sampling or hashtag/key research) is only valid for the part of television households who use social media for communication.

The search for a systematic way to approach social media data will be in the focus of this work.

4.1 Television audience and microblogging

"At first sight, microblogging and television inhabit different worlds" (Bredl et al., 2014, p. 196). Twitter created the opportunity for TV audiences to comment on their viewing experiences and to share opinions and "among the most prominent uses of Twitter is its role in the discussion of widely televised events" (Highfield, Harrington and Bruns, 2013a, p. 315).

Twitter is able to act as a backchannel for mainstream television (Deller, 2011; Harrington, Highfield, Bruns, 2013b; Bruns and Stieglitz, 2012). Bredl, et al., discuss microblogging as a new approach to audience research as "Twitter seems to offer great opportunities for researching audience conversations about television" (2014, p. 201). Webster (2010) describes Twitter "as an accompaniment to live TV", as Twitter users in the United States regularly spend more than three hours a day watching television. Furthermore, 50% of the popular US television shows (with high Nielsen television ratings), are ranked very high within Twitter trends. "Ever since the notion of an 'active' audience was firmly entrenched in media studies several decades ago, it has been recognised as a medium that readily catalyses audience discussion, interaction, fandom and other social activity" (Harrington, Highfield and Bruns 2013b, p. 405).

According to Twitter (2014) the conversation on Twitter about live television in the United States is growing and they report that 19 million unique people in the United States posted 263 million tweets about live television in Q2 2013. They furthermore reveal that on average the number of people reading the Tweet is fifty times larger than the number of Tweets. Bredl et al. (2014) cite Kepplinger and Martin, Keppler, and Ulmer and Bergman who explored how families use television to interact and they found that media content is discussed in the majority of public conversations every day. As Twitter connects television audiences it therefore becomes a kind of "virtual lounge room" where the show is watched virtually together, the equivalent of people who gather around the watercooler in the office and talk, metaphorically a "watercooler in the cloud, but one where the watercooler conversations take

place instantly, rather than at work the following morning" (Harrington, Highfield, and Bruns 2013b, p. 405).

Twitter research reveals information which cannot be provided by traditional audience ratings measurements (Harrington, Highfield and Bruns, 2013b). Social media monitoring and opinion mining for the television industry tracks and analyses "how viewers engage with programs, how they feel about the programs (and advertisements) they consume and how they respond to them" (Napoli, 2012, p. 9). "Social media have emerged as the primary means by which audience engagement with television programs is being translated into performance metrics that can be used in the assessment and evaluation of program audiences" (p. 9).

Hightfield, Harrington and Bruns (2013a) investigated the interconnection between Twitter and selected television events and conducted analysis to see whether or not a pattern of Twitter activity can be found related to the television show. Users mark their messages by including keywords and topics which enable users to search for discussions and messages related to particular topics (Züll, Mikelić Preradović and Boras, 2013, p. 279).

Television channels include hashtags which are related to major media events or daily or weekly television shows (Bruns and Stieglitz, 2012) so that audiences can identify the discussion about programs on Twitter and thus create the opportunity to participate in discussion in the context of their television consumption.

Torrez-Riley (2011) describes microblogging as a trend and technology which enhances the social television experience. In an era of fragmentation, digitisation and increasingly convergent digital media usage, Twitter is being used as a tool to comment and share television experiences (2011). Research during recent decades discovered that the discussion of media content dominates the everyday public communication (Kepplinger and Martin 1986; Keppler, 1994; Ulmer and Bermgann, 1993). Buschow, Schneider and Ueberheide investigated categories of communication activities during television reception in Germany (2014).

The comments and opinions about the watched content expressed within them, and metadata, such as when (time) and where (location) comments were posted are of high interest for the audience researchers as they provide huge amount of data about the television audiences - even though the data is not statistically representative. The numbers of comments related to television consumption are high, especially for global mass television events and the amount of

data available provides reliable insights in addition to traditional, sample-based television rating-measurement.

Given the situation regarding how data can be used and accessed, Twitter is of particular interest for researchers, as it provides the most open access to data via the Twitter streaming service. This applies to Twitter as a global network worldwide and is the most important advantage compared to Facebook. Twitter makes data available almost in real-time to the researcher through access to the data stream.

Nevertheless, compared to the wider television audience, microblogging remains a niche phenomenon in Europe and the online activities of users are not always connected to television (Buschow, Schneider and Ueberheide, 2014).

4.2 Twitter: structure and communication

Twitter was founded in 2006 and with over 270 million active users and 500 million posted tweets per day, it is the most popular microblogging service in the world (Twitter, 2013). As a worldwide network Twitter supports more than 35 languages and 77% of Twitter accounts are outside the United States.

Various authors describe and summarise the functionality of Twitter in a comprehensive matter (Bruns and Stieglitz, 2012; Bredl et al. 2014).

Twitter enables users to communicate and interact in real time with each other. Users send 140-character comments (messages) called “tweets” about various topics and post (publish) these comments, which can be any kind of personal or business information. Other users follow these messages (“followers”), some retweet, for example when they agree with the content. These short messages are truly advantageous for the researcher, as they are easier to analyse thanks to their compact nature and are frequently equipped with emoticons which are significant for opinion mining. The popularity of forwarding and sharing links from a tweet to another extends the interaction to a broader user ship. The user either participates in an already ongoing conversation or is able to initiate a discussion. Furthermore, the user can participate as a follower, which means that the user follows the communication of others. “Although Twitter provides a customizable profile page, the key principle is to ‘follow’ the messages of other Twitter users” (Bredl et al., 2014, p. 198).

One main typical practice on Twitter is the usage of hashtags (Bruns and Stieglitz, 2012; Bredl et al., 2014). Hashtags are keywords which are prefixed by the hash symbol #. Bruns, Highfield and Harrington (2012) explain the functionality of Twitter as a virtual lounge room, where people virtually watch together.

The Twitter user can be an individual or an organisation, it can be used for private or, for example, for marketing reasons to share and spread information. The mechanism of retweeting has been adapted by the user to disseminate information: retweeting means that the user shares information (with their followers) and may also add a positive or negative comment (Bruns and Stieglitz, 2012). With the possibility to retweet, Twitter has become a viral media, where information can be spread quickly and extensively. The power of Twitter is its potential for mass distribution of information as "electronic word of mouth" (Jansen, et al., 2009, p. 2169).

As the majority of Twitter users are mobile users, it plays an increasingly important part as a platform for public communication in various topics of publicly shared interest.

Bruns and Stieglitz (2012) identified 3 different types of discussions which can be observed on Twitter:

- **Politics**

  Twitter is often researched in the context of elections and as a tool of political parties used during an election campaign (Bruns and Highfield, 2013). Some authors demonstrate that it is possible to use Twitter analysis to a certain extent to forecast election results. It can predict the opinions of unknown people on a certain topic (e.g. result of a referendum or elections) and was already used to forecast the outcome of a forthcoming election in different countries, such as Hungary or Italy (Berend and Farkas, 2008; Artemi, 2005).

  Twitter is not always popular among politicians especially in countries with authoritarian or totalitarian governments, as could be seen in particular during the Arab Spring (Züll, 2014). Bruns and Stieglitz (2012) refer to a publication of Stieglitz and Dang-Xuan (2012) which provides a general framework for social media analytics in the political context.

- **Natural crises and human disasters**

  Twitter is often researched in the context of crisis communication and diffusion of communication. As Twitter is a high speed communication tool, many of these studies point to the important role Twitter has played in quickly disseminating information about world events (Acar and Muraki, 2011).

- **Entertainment and communication related to brands**

  The type of discussion includes communication around brands (e.g. consumer goods) and big entertainment formats such as television shows. With the active participation of television viewers in social networking and microblogging services such as Twitter, the communication about their viewing experiences is recorded in digital form and can be found on the internet.

"One particularly interesting phenomenon is the use of Twitter to connect and support conversations between audience members for live or mediated entertainment" (Harrington, Highfield and Bruns 2013a, p. 315).

In contrast to other social networking sites, in particular Facebook, content on Twitter is not limited to a group of "friends" but is completely open to the public and it is therefore possible to locate communication or to track the Twitter stream.

Twitter ranks internationally and in Croatia (figure 3.3) behind Facebook. Bredl, et al., cited Herwig who stated that "Microblogging can be considered as a form of blogging" (2014) as Twitter enables users to communicate and interact in real time with each other.

Figure 4.1. Ranking of social media site in Croatia (Alexa, 2016).

The top social media sites in Croatia (Alexa, 2016) are shown in figure 4.1. Extracting the social media sites in Croatia (figure 3.3) according to the definition of Haenlein and Kaplan (2010) search engines such as Google and Croatian news portals such as Index.hr, Net.hr, Jutarnji.hr, and 24sata.hr are excluded, one can find Twitter among the top five most popular social media sites. Facebook is the number one social media site with 1.2 million users (2016) in Croatia, followed by the video file sharing website YouTube (Alexa, 2016). After Wikipedia, Twitter is in position four. Finally, Blogspot.hr holds the fifth most popular website position.

Although in Croatia the microblogging service is nowhere near as popular as in the United States or in the United Kingdom, its popularity is growing fast: in 2013 the Croatian Financial Times (Runje, 2013) reported that the number of Twitter users in Croatia had reached 51,986 and predicted that the popularity of Twitter in Croatia would grow (figure 4.2). The service has been increases among a number of demographic groups and urbanites (compare Duggan, et. al., 2015).

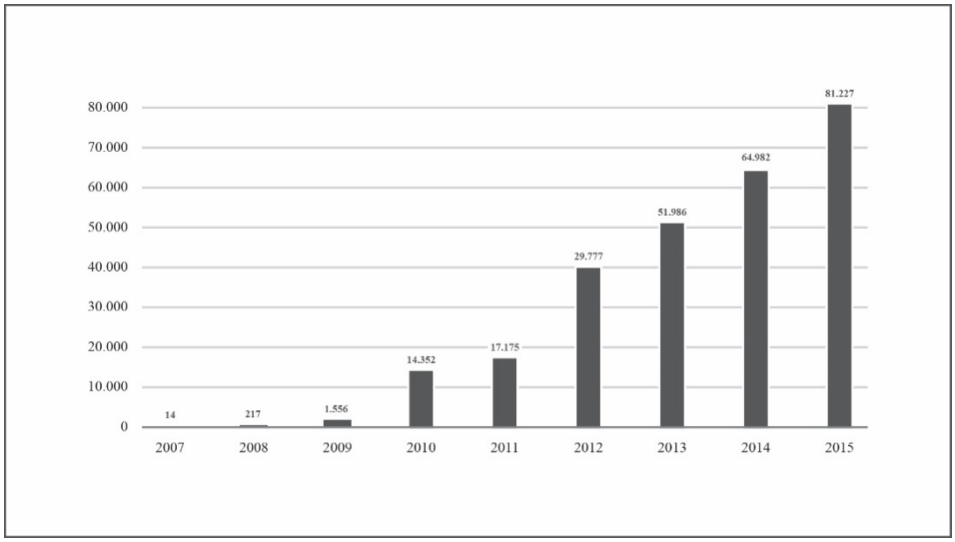


Figure 4.2. Twitter usage numbers in Croatia (Hrastovčak, 2013).

Many of the issues discussed on Twitter are related to television programs (Deller, 2011; Wohn and Na, 2011). As the use of Twitter has become more commonplace throughout the world, its role in politics and governments has also increased. The Croatian government used Twitter during huge floods which occurred in the spring of 2014 and for almost three weeks constantly tweeted about the situation in the affected areas.

"Croatia, particularly the Government, is very active on Twitter, currently counting over 60,000 followers, 42 tweets per day and increasing rapidly. In addition, we have a strong presence via the Ministry of Foreign and European Affairs, EU Commissioner Mimica, two heads of missions and several spokespersons in ministries" (Government of the Republic of Croatia, 2015).

In 2015 the government of Croatia was named the most prolific government on Twitter in Europe and the 2nd most communicative government in the world (Government of the Republic of Croatia, 2015):

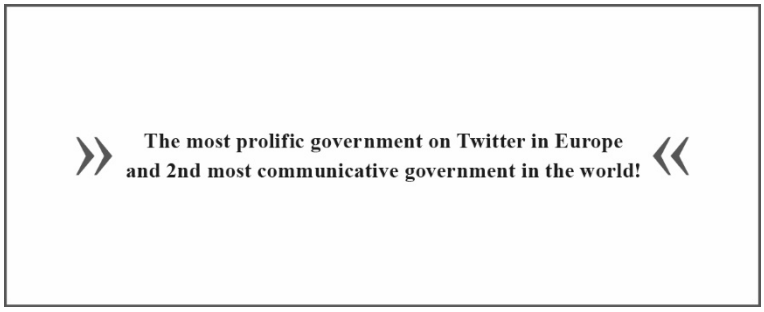


Figure 4.4. New Twiplomacy Study voted (Government of the Republic of Croatia, 2013).

The government of Croatia appreciates the two way communication to connect with audience and influencers – to create lists and to engage and call followers to action and on the Central Government Portal (Government of the Republic of Croatia, 2015) they remind users to include the at sign @, and to not forget the hashtag sign #.

Digital technology and the internet as global memory provide society with a large data archive and enable the world of researchers and scientists to use this archive for analysis. The Twitter data sets are large and continuously growing so traditional data processing often reaches limits. "Big data" is, in many ways, an insufficient term. Big data is less about data that is big than about the infrastructural capacity to search, aggregate and retrieve results (Boyd and Crawford, 2014). In April 2010, the Library of Congress, the largest library in the world, located in the United States and Twitter signed an agreement that Twitter would donate all tweets from the inception of Twitter up until the date of the agreement (Library of Congress, 2013). According to this agreement (appendix 3) by 2013 the library had already achieved an archive of 170 billion tweets, totalling 133.2 terabytes for two compressed copies.

The aim of this agreement is to make the archive accessible to researchers and policymakers, creating an infrastructure for the organisation of the big data. "As society turns to social media as a primary method of communication and creative expression, social media is supplementing and in some cases supplanting letters, journals, serial publications and other sources routinely collected by research libraries. Archiving and preserving outlets such as Twitter will enable future researchers access to a fuller picture of today's cultural norms, dialogue, trends and events to inform scholarship, the legislative process, new works of authorship, education and other purposes" (Library of Congress, 2013).

The agreement emphasised that there are still technology challenges to work around to allow the researcher access to large data sets, as the Twitter archive is a new type of digital

collection, in particular as future tweets come into the library in a continuous real-time stream. The main challenge for the library is to develop a stable infrastructure and sustainable way of organising the massive collection, although the library mentioned that "the technical infrastructure for the Library`s Twitter archive follows the same general practices for monitoring other digital collection data at the Library" (Library of Congress, 2013).

The technology to "allow for scholarship access to large data sets is not nearly as advanced as the technology for creating and distributing the data. Even the private sector has not yet implemented cost-effective commercial solutions because of the complexity and resource requirements of such a task" (Library of Congress, 2013). The cooperation between the Library of Congress and Twitter indicates the shift of libraries and research towards a digital future, and the question is whether other libraries will follow or if Twitter itself opens its archive to the researcher, which they sometimes very selectively do.

> "With more than 500 million Tweets a day, Twitter has an expansive set of data from which we can glean insights and learn about a variety of topics, from health-related information such as when and where the flu may hit to global events like ringing in the New Year. To date, it has been challenging for researchers outside the company who are tackling big questions to collaborate with us to access our public, historical data. Our Data Grants program aims to change that by connecting research institutions and academics with the data they need" (Twitter, 2014).

They invited applications for this grant for special researcher purposes which meant getting access to data for a certain timeframe for selected hashtags to build up a data collection (#Data grants pilot program), as was performed as part of this research (Chapter 5; appendix 1). Social media data as library data may allow for the interpretation of social media data as a public good (Bergstrom, Blume and Varian, 1986) in a public library from which individuals and researchers cannot be excluded from accessing. What kind of analysis could be performed if the television audience analyst could have free access to a complete data collection? The

Twitter data grant pilot program inspired an approach to the topic under the assumption that it is possible to have access to data collections.

Twitter is known among scientists for its more open data policy although changes are expected (Kern, 2012). Using the API (Application Programming Interface) of Twitter it is possible to capture data sets of the tweets although there are limits in respect to the API and the reliable real-time Twitter analysis tools (Bruns and Stieglitz, 2013; Bruns& Liang, 2013).

The researcher cannot control the Twitter API (Courtois and Mechant, 2014) and a conclusion with certainty regarding the composition of data cannot be reached. There is a lack of transparency but it is used by the researchers within its limits (Bruns & Liang 2012). According to Kennedy (2012), the key concerns relate to accuracy of analysis, and quantity and 'cleanliness' of data. Twitter explains the limitation of the Twitter search API on its developer page (Twitter Developer, 2012): "it is important to know that the Search API is focused on relevance and not on completeness. This means that some Tweets and users may be missing from search results".

4.3 Approaching Twitter data for audience research

Twitter constitutes an archive of communication data. Researchers can analyse an immense amount of data by designing an appropriate program or algorithm and in addition, the internet provides open source solutions such as *social analysis tools*, which are not typically used for research but presented and used, although the availability of web applications is changing dynamically (Bredl, et al., 2014). Tools presented in this work offer the possibility of monitoring social media. Twitter-oriented websites generally allow the user to "take the pulse" of particular topics on Twitter. Consumers can evaluate products, companies can research public opinion about products, and governments can use them to create condensed versions of voting opinions.

The mining of public opinion is a process of automatic recognition and extraction of opinions or sentiments from the unstructured text. It aims to identify positive and negative opinions, emotions and evaluations (Wilson, Wiebe and Hoffmann, 2005). Collecting data and archive web data enables researchers to trace and examine how various actors interact and communicate.

The building of data corpuses by collecting data, although called web archiving, is a "particular useful method in studies of users' communicative practices" (Lomberg, 2012, p. 220). Internet research often faces with unstable objects of study (Rogers, 2013a). Archiving internet data can stabilise and preserve the research. Lomberg describes web archiving as a method that allows the researcher to retrieve and harvest internet data, which means twitter texts and relevant meta data (e.g. time stamp, geolocation or relevant user profile information). Besides the quantitative studies around political or other types of events (Bruns and Burgess, 2011) there are also qualitative studies addressing key questions of audience research, for instance concerning opinions of television audiences.

Web archiving can be defined as the archiving of internet data as well as the use of digital research tools (Brügger, 2005). Web archiving originates in public national and international institutions, for instance the Library of Congress, that serve to preserve internet and social media data as public heritage, e.g. Twitter data (2013, appendix 3).

Archiving internet data in general and social media data (e.g. Twitter data) in particular, poses significant questions in respect to capturing, archiving, structuring, and organising data. Brügger (2011; Lomborg, 2012) distinguishes three different strategies of harvesting and archiving data of the web. The *snapshot* strategy, were a certain portion of data, often used to harvest data from a large number of websites is archived (Lomborg, 2012).

According to Lomborg (2012) this method is not suited to a longitudinal study and furthermore it does not capture depth (communication level of the website). The *event* strategy involves harvesting data in relation to an event (e.g. television event), thereby collecting data for a thematic archive. According to Lomborg, this strategy suits examining communication structures on the internet and most existing social media research has used this strategy of web archiving typically using hashtag/keyword searches (Bruns and Burgess, 2011; Bedl, et al., 2014; Bruns and Stieglitz, 2012). Bruns and Stieglitz (2012) described how to retrieve twitter data sets using the open-source software *Yourtwapperkeeper*.

Besides opinion mining tools the application of open-source software is not typically used for research as they develop dynamically and offers are in continuous change. This is indicated as well by the disappearance of some programs or transformation in commercial offers.

Nevertheless, following Bruns and Stieglitz (2012) and Bruns and Liang (2012), *Yourtwapperkeeper* is utilised to capture data through the Twitter API.

As opinion mining and data retrieval based on tools such as *Yourtwapperkeeper* are based on keyword and hashtag searches it remains a challenge to select them the moment they appear on Twitter.

In case users introduce their own keywords or hashtags it has to be assumed that the captured data stream cannot guarantee a comprehensive Twitter data set. However, in this part of the empirical study it will be of relevance to analyse the data set which can be retrieved in its entirety, to see which insights can be gained which could be useful for television audience research. Bruns and Stieglitz (2014) conclude that Twitter activities around defined themes and events are not random and that patterns in particular for mainstream events can be found.

The open software *Yourtwapperkeeper* enables the researcher to create a tool which can be used to perform selected queries and automatically archive tweets which contain selected hashtags and keywords using the Twitter streaming API (Bruns, 2011). *Yourtwapperkeeper* as open-source software has to be installed before it can be used by the researcher (Bruns, Highfield and Harrington, 2014). Researchers are required to run it on their own web servers, and it should not be used for the public sharing of data sets (Bruns, 2011). After the installation of the tool, queries can be defined and it is possible to archive and retrieve data over a longer time period. *Yourtwapperkeeper* only provides post hoc data which means that tweets are captured for set keywords as they are sent (Bruns and Stieglitz, 2012). To learn more about the functionality and context of *Yourtwapperkeeper*, see Bruns and Liang (2012).

It has to be considered that the types of hashtags which need to be selected are not the ones that may be observed on Twitter, as Croatian Twitter users may use other keywords to organise their communication, given that the promotion of hashtags or keywords by mass media is still in its infancy (Tewes, 2015).

4.3.1 Opinion mining for audience research

The issue of finding opinion sources and monitoring them on the web is a challenging natural language task due to a large number of diverse sources, each containing a huge volume of opinion text hidden in long forum posts and blogs. The main problem of automated opinion discovery is to find relevant sources, extract opinions and organise them into usable forms (Liu, 2010).

Opinion statements are often more subtle than objective ones, sensitive to context and sometimes ironic (Lula and Wojcik, 2011). Although the importance of the topic was recognised quite early in scientific literature (Carbonell, 1979), a widespread research interest in opinion mining developed only at the beginning of the current century (Das and Chen 2001; Tateishi, Ishiguro and Fukushima, 2001; Tong, 2001; Dini and Mazzini 2002; Pang, Lee, and Vaithyanathan 2002, Turney, 2002; Morinaga, et al., 2002; Dave, Lawrence, Pennock and 2003; Nasukawa and Yi, 2003.; Liu, Lieberman and Selker, 2003; Wiebe, et al., 2003; Cardie, et al., 2003; Yu and Hatzivassiloglou, 2003). Opinion-oriented information-seeking systems deliver the "computational treatment of opinion, sentiment, and subjectivity in text" (Pang and Lee, 2008, p. 1).

Opinion mining of any text written in any natural language employs various statistical and linguistic approaches. In a statistical approach to opinion mining, machine learning algorithms, a branch of artificial intelligence, analyse whether keywords in a text have a positive or negative sentiment based on the binary analysis, focusing either on the frequency of words or relevance of co-occurring words. Linguistic approaches use opinions libraries (Panian, 2010) containing thousands of words that represent attitudes or preferences that indicate an opinion about a topic or they utilise the publicly available resources such as Wordnet-Affect (Strapparava and Valitutti, 2004), SentiWordNet (Baccianella, Esuli and Sebastiani, 2010) and SenticNet (Cambria, Havasi and Hussain, 2012) to extract the information. Also, classifiers at the document level (Dzitac and Moisil, 2008) are used for sentiment classification.

Through a mixed-method approach, the public monitoring tools enable the identification of opinions and quantitative changes of Croatian tweet volume in a selected time frame (around the event).

Some currently active opinion mining and social monitoring are presented and evaluated as examples of real-time opinion mining in respect to the information they provide to the researcher (Züll and Mikelić Preradović, 2013):

- Sentiment140
- Tweetfeel
- Tweettone
- Topsy
- Social Mention

The tools were evaluated in respect to the provided information and its functionalities. The openness of the tool providers turned out to be one of the main quality factors, as it enables the researcher to evaluate if the tool offers a suitable research framework and reliable information.

The potential of opinion mining tools is demonstrated in selected examples of television programs to show the usefulness of social media analysis for television audience research. It is necessary to select television programs of diverse character (e.g. serial or non-serial television programs). Within the framework of this research work, three strong television brands were used as an example for comparing and testing the results of the tools. The question is if the television broadcaster provides information about the program on social media with the aim to promote the program and to motivate users to communicate about it. This could be considered as a precondition, to create a relationship between the audience and the program. Tools for mining opinions generally work in a similar way to search engines and as a precondition for the analysis the selection and definition of queries is necessary.

This can be considered the crucial point of this research as one needs to assume that Twitter users use the abbreviation or hashtag which is, for example, promoted by the broadcasters to discuss the television program. Therefore, in order to use the tools for analysis it needs to be a precondition that users are aware and open to using hashtags and keywords proposed by the broadcaster and furthermore broadcasts need to promote these. Usually, one or more single keywords, key-terms (brand) or hashtags which can be entered in a search field of the individual opinion mining tool are selected.

The applications collect the most recent tweets which will be submitted as the result of the query. Changes in the quantity of tweets in a selected timeframe will raise questions regarding whether or not this could be related to the content in the broadcast and compared with television ratings and the audience flow. It could be interesting to investigate if Twitter usage related to television content shows similar patterns compared to television usage. Qualitative examination of tweets and mining of opinions could depict television audience opinions to better understand what television viewers like.

The above mentioned opinion mining tools are presented and used in parallel to test if social media activity towards a selected television program can be measured and what kind of data sets these tools (applications) offer (chapter 5, empirical study). The measurement of social activity with several tools allows the evaluation of the tool itself. The openness and

transparency in respect to the given information and data sets can be interpreted as a quality factor for the selected tool and enables the researcher to find out if the selected tool offers a suitable framework for the research task.

The empirical part will reveal if any of the tools tested within the framework of this study are helpful for finding opinions of television audiences, to measure the social resonance and pulse of selected television programs.

4.3.2 Twitter web archives (collection and corpus)

It would be very attractive to the researcher in Croatia to have access to the Croatian Twitter library, as is the case with the US Library of Congress, known as the Twitter Research Access project (2013; appendix 3). The US library has published that the Twitter archive could provide future generations with invaluable records of communication data from the 21st century.

The project enables researchers to use the collection to study for insights into historical events, from elections to natural disasters and cannot be used for commercial purposes, to avoid interference with Twitter's core business of selling data for commercial market research.

In Croatia, like in other European countries, no agreement between state libraries and Twitter has been planned yet. As the project in the United States is ambitious and researchers there are still waiting, libraries in Europe are in a waiting position.

The management of the project is a technological challenge (Scola, 2015). First of all, the library must organise and store a static archive of tweets from the start of Twitter in 2006 to the signing of the agreement and it must also take in regular updates that number around 400 million tweets a day. The main task for library experts in technology, research and library sciences is to find a way to index the tweets and make them searchable by the researcher.

A Croatian Twitter archive would enable future Croatian researchers access to a fuller picture of today's cultural norms, dialogue, trends and events to inform scholarship, the legislative process, new works of authorship, education and other purposes. Social media researchers all over the world are waiting to see if the Twitter libraries will become a source for historians in the future.

The agreement of the library of congress shows that Twitter data could be considered s public good (Bergstrom, Blume and Varian, 1986) which should be accessible by public

institutions. A question that arises is whether or not European governments should develop a common position towards social networks to agree on scholarship projects and access to data.

Lomborg (2011) focuses on selective archiving strategy although it is not as prominent in social media research. It can be used to focus on a small number of preselected websites or user accounts. Regardless of archiving strategy, the method of collecting data for analysis creates a number of challenges. No real-time data collection is guaranteed to be entirely comprehensive if hashtags are not prominent enough or not foreseeable. The "ideal scenario" for a social media analyst would be having access to a huge data collection of social media data similar to the Twitter digital data collection of the Library of Congress (4.2) for the United States or through a Twitter data grant (4.3 and appendix 2) which was applied for as part of this research (appendix 2). Brügger (2011) and Lomberg (2014) explained the selective archive strategy in the context of selection of websites or accounts. Other selection criteria for data retrieve are also applicable.

According to the selected criteria, a certain amount of Twitter data sets can be collected and archived. It is necessary to define characteristics in the production of data to define appropriate boundaries such as selected geographic area, selected language and time frame, which defines the size of data volume (pulled stream) of the Twitter web archive.

The size of the data collection is influenced by factors such as

- the purpose and timeline of study
- selected regions (countries/geographic areas),
- technical capacity: server & database infrastructure
- research capacity: human resources
- language: language such as English cover more geographic areas, local differentiation based on language is therefore not always possible

Ljubešić, Fišer and Erjavec (2014) used the language as selection criteria. The aim of their analysis was to build a Croatian, Serbian and Slovene web corpus, using a method with seed terms and language identification modules to build the web archive, the so-called data corpus. Full records of Twitter data and the compiled corpora formed the empirical basis of the

study (5.2) as an approach to finding Twitter data without using hashtag/keywords search. A number of fundamental challenges require considerations; e.g. the missing differentiation between Croatian and Serbian Twitter users could be considered a limitation.

For social media analysis, web archiving is useful for creating a stable researchable corpus with relevant Meta data (Lomberg, 2014). The key concept of traditional audience research is based on sampling (figure 3.6), which means a selection of subsets of individuals from within a statistical population, which is in this case the web corpora, to estimate the characteristics of the whole data collection and to systematically study Twitter communication that would have been almost not possible to do for Croatia. Usually, such content analysis based on drawing representative samples is analysed to describe typical patterns or characteristics or to identify important relationships (Riff, Lacy and Fico, 2014).

While the volume of the data collections increases, so does the complexity. The focus is on finding best feature values to represent each observation, which is similar to using a number of data fields, such as age, gender, income, education background etc. known from traditional audience analysis, to characterise each individual data set. This type of sample-feature representation inherently treats each individual (in this case a Twitter data set or Twitter profile) as an independent entity without considering their social connections (Wu, Zhu, Wu and Ding, 2014). "Such social connections commonly exist in not only our daily activities, but also are very popular in virtual worlds" (Wu, Zhu, Wu and Ding, 2014, p. 102). It is worth mentioning in this context that traditional audience research and television audience measurement does not take into consideration social influences among family members in panel households (chapter 3.2) while watching television.

Web archiving can complement the analysis of large data sets, enabling the analysis and detection of topics and discussions, and can also provide some qualitative exploration (Highfield, 2012). While this analysis focuses in particular on the finding of media related keywords or hashtags and comments on a particular issue, implied connections among users and topics (topical networks) may also be used to examine different kinds of communication (Highfield, 2012). The correlations between individuals inherently complicate the whole data representation and any reasoning process; however the sample-feature representation is used, where individuals are regarded similarly if they share similar feature values (Wu, Zhu, Wu and Ding, 2014).

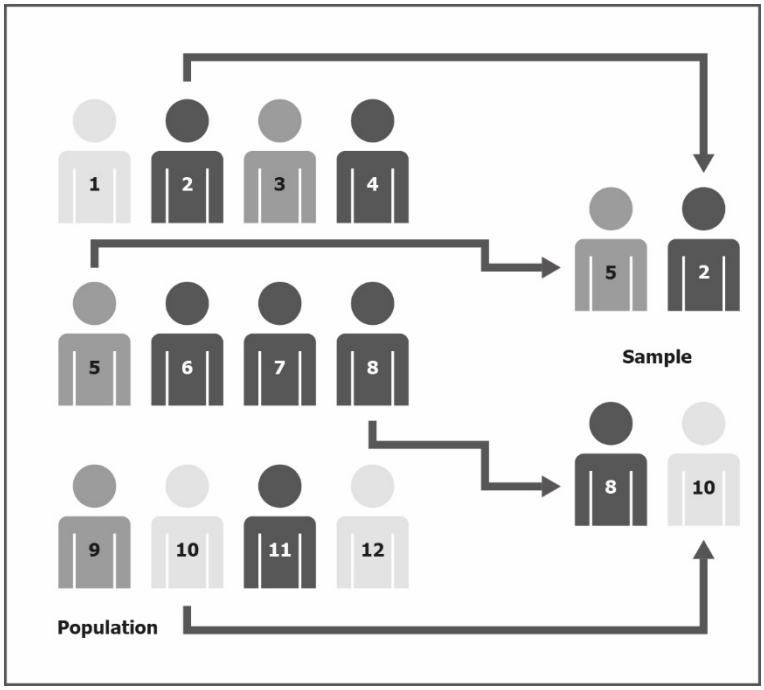


Figure 3.6. Sample-feature representation (Cochran, 1977; Wu, Zhu, Wu and Ding, 2014).

The sample-feature representation is an early stage of the analysis of data systems and comprises several steps:

- Defining the Twitter web archive (data collection) of concern as a population
- Specifying a sampling frame, a set of items, hashtags or possible events to measure
- Specifying a sampling method for selecting items or events from the frame
- Determining the sample size
- Implementing the sampling plan
- Sampling and data collecting
- Data which can be selected

The assumption which had to be proved in the context of this work was if it is possible to find television related tweets (and keywords for a further keyword search) in samples which could be considered representative for the population of data.

The proposal of this research is to take samples of a data corpus (defined data collection) and to identify television related tweets and keywords.

It will be experimented if it will be possible to identify trending media related topics (comments) on Twitter (Benhardus, J. and Kalita, 2013) within the frame of the empirical research of this work. Following the identification of relevant comments, a series of quantitative and qualitative tasks (Berger, 2015) are applied to analyse large data sets in greater detail. The first research task is to identify and evaluate possibilities of getting access to a large Twitter data collection which could be considered as representative for the research topic. In the frame of this work different approaches to data collections were discovered.

Sampling on a Twitter data corpus as the second research task submits data relevant to a specific topic in which a predetermined number of observations are taken from a larger population. The proposal in research step three is to analyse samples to find television related key terms. It is expected that key terms related to television will be found – and this is of particular relevance for this research – key terms which are not mandatorily the title of the show, or promoted by the television broadcaster, but key terms which may be introduced by the audiences themselves. After extraction of keywords (research task 3) search queries can be defined (research task 4) to search in the defined corpus and/or search real-time tweets using existing opinion mining tools (4.3.1).

Table 4.2. Twitter data collection processing framework.

1.	Evaluation of possibilities to create or get access to a Croatian Twitter data corpus
2.	Sampling on Twitter data corpus and identification of tweets related to television
3.	Key word mining: Extraction and ranking of TV related key words
4.	Search queries based on pre-defined list and finding on TV related tweets in the defined corpus and/or real-time

The method to work with Twitter data collections and sampling seems to be suitable for research as the object is stable and can be approached without the precondition to know media related keywords. The analysis could be repeated and therefore an ideal research environment can be created for public research.

Ethical aspects need to be taken into consideration for ongoing ethics discussions in internet research.

This study introduces sampling of tweets as an approach to finding television related keywords which are used in Twitter conversations. The aim is to first find a statistically relevant number of television related keywords in the Twitter archive which can be used for further research (Oreščanin, 2014). The method of sampling is well known in traditional television audience research.

5 EMPIRICAL STUDY

The aim of the empirical study is to, in a systematic and controlled manner, understand and explain data which is created by human activity on social media, in particular on Twitter. It is important that the social network is accepted by the public and that increased social media usage can be expected. Whether or not an approach to this data can be found and explained in Croatia as an example of a television market in Europe shall be investigated.

The aim is to examine the Twitter activity of television audiences. It is also an aim to learn if and how data created by television audiences can be located, and if the data could submit information which one could interpret to complement existing television audience analysis. Additionally, the character of social media data shall be investigated to reveal if the information, which is provided with the tweet itself, such as time of creation of the tweet, or a geographic reference, could be helpful to understanding the behaviour of television audiences.

Ethical aspects have to be taken into consideration (Berry, 2004) and the data which can be found on the internet has to be screened to protect the internet user although the data is (partly freely accessible) to the researcher (Züll, 2013; Berry 2004). As Twitter data sets reveal information about the users, such as location, IP address, and profile pictures, it is justified that the data can be used for analysis, and the way that it is used for analysis is also justified. However the social media user should be aware that open communication on the internet can be compared to speaking in front of an audience, and that comments on social media networks can be equated with an individual release of information. However, portraying information in a way that could be connected to an individual person has been consciously avoided.

Given that people use social media platforms and social media data can be located on the internet, it is proposed that Twitter data be analysed to learn if audiences' comments and opinions on social media can be used to overcome the limitation of conventional television ratings (Wakamiya, Lee and Sumiya, 2011). In the empirical part the exploratory surveys of this study will be described, conducted using Twitter messages targeting television programs in Croatia.

There is also a particular focus on the search of opinions regarding selected television programs, as this information is only indirectly submitted by television audience research because people-meter-methodology only submits quantitative aspects of television usage. It will be investigated whether or not Croatian television viewers comment on Twitter and if and

how these comments can be identified. As many open-source opinion mining tools exist, the research will start with determining if these are suitable for analysis. To understand the data and information structure of Twitter data, open-source software was applied as an example of a data retrieval method.

The market of social media analysis tools is in continuous change and development, therefore this part should be taken as a general example to demonstrate how Twitter data sets can be retrieved and analysed. The second part of the research focuses on the problem of finding a fundamental solution to treating Twitter data for research and the US solution of interpreting Twitter data as a data library which should be accessible by (public) researchers. Working with a database (data corpus) enables the research to have a broader basis, compared to the limited possibilities of opinion mining tools.

5.1 Discovering social television response using opinion mining tools

The scope of opinion mining is to identify opinions and sentiments regarding selected topics, in this case the selected television shows. The results of this part of the research were published in 2013 in WSEAS Press (Züll and Mikelić Preradović, 2013).

Opinion mining or sentiment analysis is a process of automatic recognition and extraction of opinions or sentiments from the unstructured text. It aims to identify positive and negative opinions, emotions and evaluations (Wilson, Wiebe and Hoffmann, 2005). As a result, the volume of comments and other comment-related aspects (metadata) became the focus of interest. An analysis of Twitter may, therefore, give insights into why particular television events resonate with the audience (Thelwall, Buckley and Paltoglou, 2011).

In general, opinion mining tools work in a similar way to search engines. Key words must be chosen, and queries must be defined, therefore a single keyword, key-term (brand) or hashtag are entered into the search field. The tools then produce the most recent tweets. Selected opinion mining tools were used to test whether or not television-related social media activities can be measured. The mining tools (applications) were also assessed for what kind of analysis they can offer: each tool was evaluated according to the information it provides to the user.

Table 5.1. Selection of opinion mining tools used for this research.

Opinion Mining Tool	Web Address	Key Term	Description
Sentiment 140	sentiment140.org	Discover the Twitter sentiment for a product or brand	search engine structure
Tweetfeel	tweetfeel.com	TweetfFeel gathers real-time Twitter data about whatever search term the user has entered	Marketing start up that combines real-time search with sentiment detection
TweetTone	tweettone.com	Real-time twitter sentiment analysis	owned by Lymbix
Topsy	topsy.com	Twitter Analytic tool	Allows estimation of number of tweets
Social Mention	socialmention.com	Real-time search and analysis	Includes other social media as source

The openness of the provider is interpreted as a characteristic which indicates the quality of the individual tool. This enables the user to determine if the application provides a suitable framework for an individual research task. The research reveals the strength and weaknesses of the tools, and reveals if any of the tools have the potential to be used by television broadcasters to learn whether or not the social resonance related to a selected television program is measurable.

There are many tools for monitoring and tracking opinions in Twitter. Sentiment140, Tweetfeel, Tweettone, Topsy and Social Mention were selected randomly for this research. They claim to measure sentiment or opinion of a user on a topic automatically. Many are free and some tools are only accessible as part of a paid subscription. Some commercial tools such as Topsy offer free trial periods for researchers, and additional filter functions and a wider scope of data sets are offered to encourage subscription.

As more data sets are provided, it is possible to extend the qualitative research by quantitative factors, such as the number of tweets. Topsy and Social Mention offer “advanced” research functionality, such as geographic or time filters. Sentiment140, Tweetfeel and Tweettone are three examples of real-time opinion mining tools, which are available to the researcher for no cost. These tools aggregate tweets about topics and provide statistics about the tweets for a selected (but limited) time frame. All of the tools were tested in the same time frame.

To test the tools, the following terms were selected: the *Oscars* (Academy Awards 2013), the American television series *American Dad*, and the international television franchise *Who Wants to be a Millionaire* broadcast on HRT in 2014 (Croatiaweek, 2014). The Academy Awards, or the official title *the Oscars*: an annual live televised red carpet awards ceremony

established in 1929 to honour professionals of the film industry. It is a one-off program with a huge amount of potential topics to talk about. The shows generally lasts an average of three and a half hours. The *Oscars* are one of the biggest television events in the world and are very much promoted worldwide. On *Oscar's* night, millions of tweets are accompanied by a worldwide public discussion. In Croatia the *Oscars* is aired on HRT.

American Dad: a popular American television series created by Seth MacFarlande, Mike Barker and Matt Weitzman. It has been nominated for numerous awards and awarded the top television series by the American Society of Composers and is broadcast worldwide. It was chosen as it was of was of particular interest to see if the tested tools were able to recognise the word *American* in the name as a part of the brand. In Croatia *American Dad* was broadcast on private television.

Who wants to be a millionaire: an international television game show franchise. The format of the show was created by David Briggs, Mike Whitehill and Steven Knight. In the show contestants play a game and a few of the participants (the top scorers) have been able to win the top prize. International variants have been broadcasted in more than 100 countries worldwide. The franchise is one of the most successful formats ever.

The *Oscars* was selected because it is a one-off program which produces a large number of potential topics to discuss. There were millions of tweets posted on the night that the *Oscars* were held. Furthermore, many tweets prior to and after the Oscar broadcast initiated worldwide public discussions about television shows, news broadcasts in which the celebrities themselves were also involved.

From a research perspective, the television show American Dad is a challenging topic: the word *American* could lead to results being produced out of context, while *Who Wants to be a Millionaire* is a long a title which could be problematic for some opinion mining tools as they might not be able to recognise the brand. The selection of the television programs was performed randomly within this research. In a wider scope of research one should develop and apply a method in the selection of television program, e.g. following the approach published in Buschow, Schneider, Ueberheide (2014), who developed criteria to select television programs for research e.g. that the programs need to be an appropriate genre for social television, that the programs only be broadcast in a specific territory, that only programs with public relevance are selected and that they have to more or less reflect the researched television market.

5.1.1 Discovering sentiment with Sentiment140

The creators of Sentiment140 (sentiment140.org) assume that the purpose of this technology is that data can be used to guage public opinion (Go, Bhayani and Huang, 2011). Sentiment140 (formerly Twitter Sentiment) enables users to research and determine the sentiment for a brand, product, or topic, based on Twitter analysis (Newmann, 2012). The system behind Sentiment140 is to apply distant supervision and training data made up of tweets containing emoticons – where the outcome is extremely generalised. The *Oscars* query delivered general results about the *Oscars* and opinions about the event, but no television related information about the event broadcast on ABC. The query performed on 26th February 2013 produced 84 positive and 38 negative tweets (69% positive/31% negative), and was generally stable over several queries (figure 5.1).

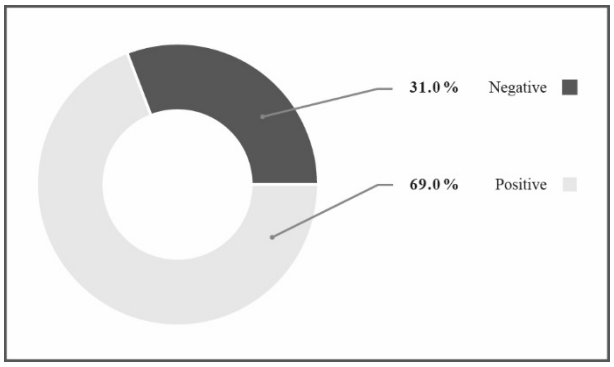


Figure 5.1. Opinion mining using Sentiment140, example: the *Oscars*.

According to television ratings, the absolute number grew as the evening approached. The modified query (with more keywords) returned disappointing results: *Oscars* AND *red carpet* delivered only 1 Tweet, while *Oscars* AND *winner/loser* returned no results. Furthermore, the *American Dad* and *Who Wants to be a Millionaire* queries proved that in general the number of keywords is not as important as the defined terms or brands. *American Dad* had five positive tweets and one negative tweet (83%/17%), while the *Who Wants to be a Millionaire* query returned 21 positive and 10 negative tweets (68%/32%).

Given that Sentiment140 only returns recent Twitter posts, it produces a small sample size, not allowing an accurate reading of online sentiment and this can easily distort the results (Newmann, 2012). Furthermore, there was no information given regarding the tweet location and source. This would be valuable for television formats like *Who Wants to be a Millionaire*,

where information about the origin country of the tweet (and therefore the country version that is the subject of the tweet) is important. The query results were stable over the course of a day, and, even though the size of samples is too small, Sentiment140 seems to be a reliable tool for analysis.

5.1.2 Twitter search with feelings: Tweetfeel

Tweetfeel (Tweetfeel.com) collects tweets about popular topics, brands and television programs, and claims to gather the best results for such queries. *Oscar's* research returned a large number of tweets and provided positive and negative comments on the topic (5.2, 5.3 and 5.4).

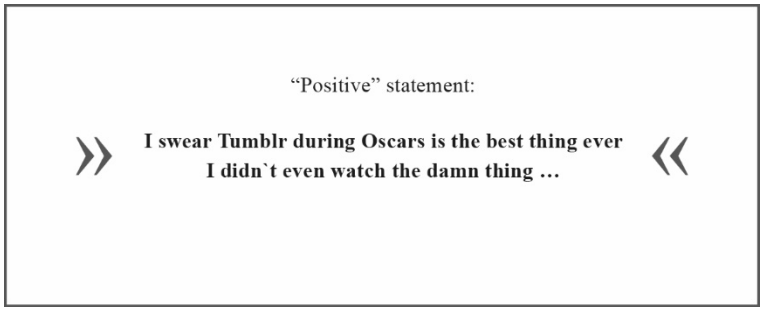


Figure 5.2. "Positive" statement towards #*Oscars* (7 April 2013).

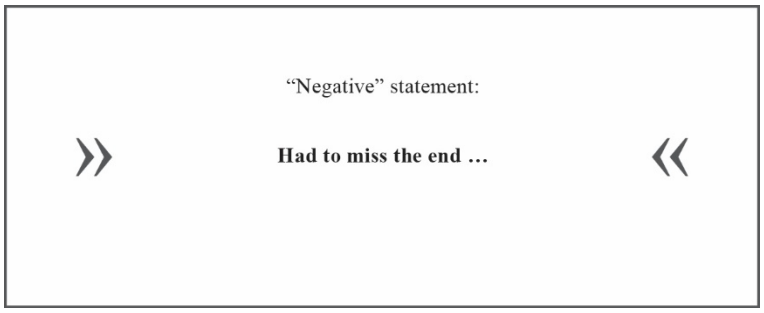


Figure 5.3 "Negative" statement towards #*Oscars* (7 April 2013).

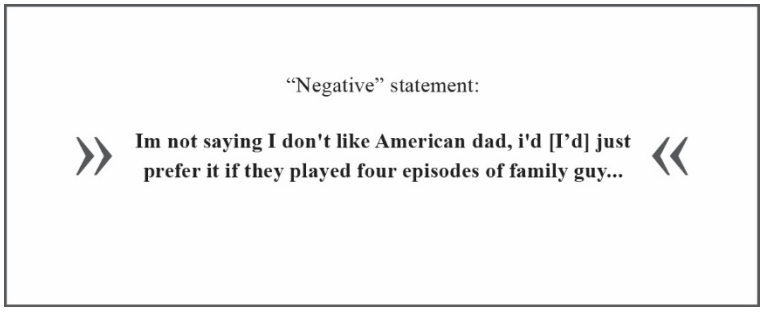


Figure 5.4 "Negative" statement #americandad (7 June 2013).

The tool does not successfully detect sarcasm or misinterpreted tweets (figure 5.2, 5.3, 5.4), as can be seen in tweets which were shown as negative tweets (this is a problem which all of the tested tools have), and it also experiences technical problems and errors with some browsers (e.g. Google Chrome), with no warning message and research does not produce results when returning to the Twitter page with the original tweet. The *Oscars* query delivered 305 positive and 281 negative tweets (52%/48%), figure 5.5, while *American Dad* had 145 positive and 125 negative tweets (53%/47%), figure 5.6.

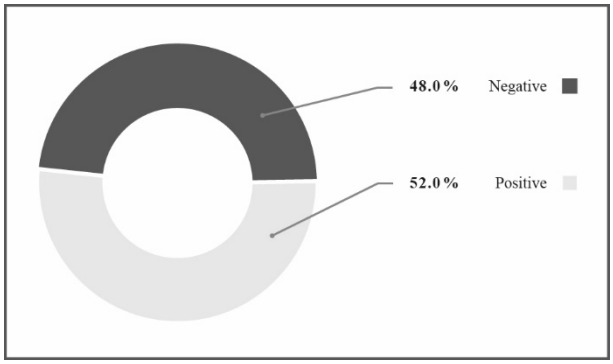


Figure 5.5. Opinion mining with Feelings: Tweetfeel, example: *Oscars*.

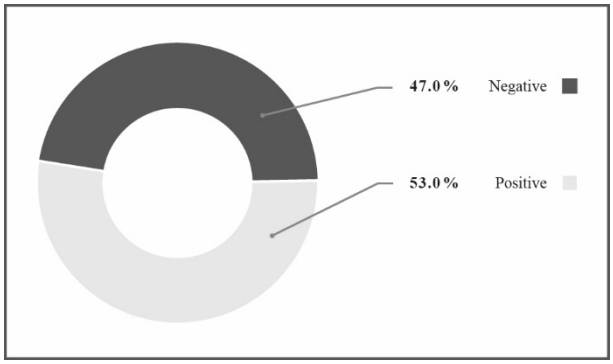


Figure 5.6. Opinion mining with feelings: Tweetfeel, example: *American Da*d.

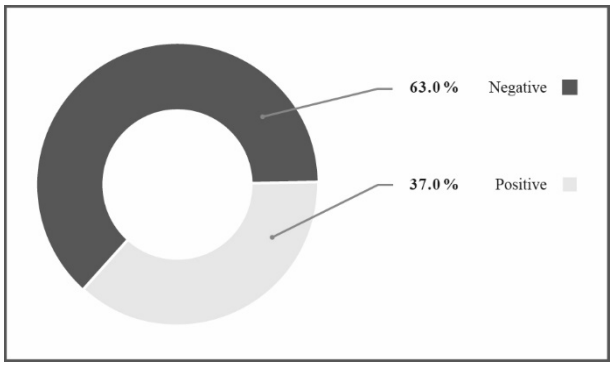


Figure 5.7. Opinion mining with feelings: Tweetfeel, example: Who *wants to be a Millionaire*.

Finally, the *Who Wants to be a Millionaire* query returned 6 positive and 10 negative tweets (37%/63%, figure 5.7). Unlike Sentiment140, the way that the tool works is not clear, and the website does not contain an explanation of its research methodology.

5.1.3 Searching the sentiment with Tweettone

The Tweettone (tweettone.com) tool categorises tweets according to eight 'emotions' (positive: friendly, enjoyment, amusing, content; and negative: sad, angry, uneasy, shame).

The *Oscars*, *American Dad* and *Who Wants to be a Millionaire* queries performed on 26th February 2013 returned insignificant results: there were 68 positive tweets and 42 negative tweets (62%/38%) for the *Oscars*, while the results for *American Dad* was unusable, because 75% of the positive tweets were posted by one user. This analysis simplified the result of this tool by summarising the tweets into positive and negative emotional categories.

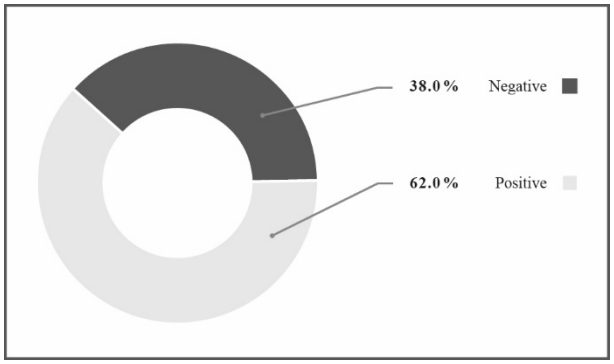


Figure 5.8. Searching for opinions with Tweettone, example: *Oscars.*

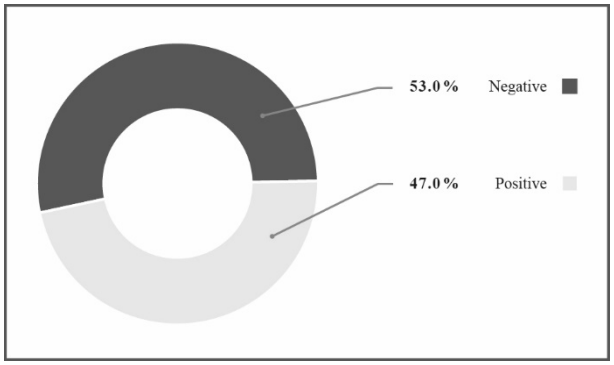


Figure 5.9. Searching opinions with Tweettone, example: Who *wants to be a Millionaire.*

The *Who Wants to be a Millionaire* query returned 119 positive and 32 negative tweets (47%/53%), figure 5.9. Details about the methodology used and the data source cannot be found on the website directly; however, information is available about Lymbix, the company behind Tweettone.

5.1.4 Getting real time insights from Topsy

Topsy (topsy.com) appeals to television media researchers as it promotes tools for researching television, providing more advanced research functionalities and filters (such as geographic and time filters) in the context of television. In addition to Twitter data, other sources such as blogs, forums, chats, reviews and other social networks are also used for data.

Topsy enables researchers "to learn from past programs" by performing "episode, story and character analysis" and compare genres in a specified time period. It offers a two-week trial period to test the advanced research search tools, time filters (older and recent), language filters, and different word queries (all words, any, and none).

The application offers history and timeline filter functions and searches can be performed according to the locations of tweets. A timeline of shared activity can be analysed figure 5.10 and the geographic distribution of tweets can also be viewed. Topsy provides information about the number of tweets, and the data is on a 30-hour rolling basis.

The *American Dad* query returned a large amount of information about social media activity. The geographic distribution could be analysed and therefore one knows in which regions the series is currently being broadcast. The peaks on the analysis indicate the broadcasting times, as viewers tend to tweet while they are watching the series (or just after watching).

Topsy claims that is draws on a range of data sources for its analysis, such as Twitter, Tumblr, Facebook and Pinterest, however only Twitter activity is shown. The *Oscars* query (figure 5.10) resulted in 4.3 million tweets: 0.8 million were positive, 0.3 million were negative, and 3.2 million were neutral tweets.

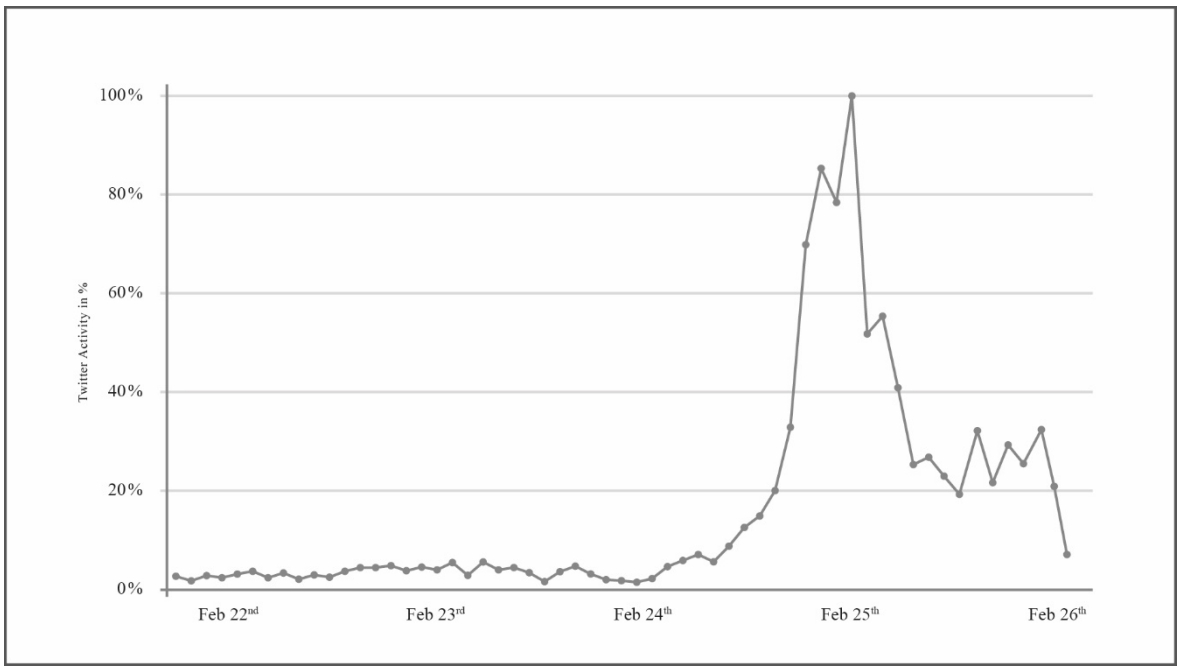


Figure 5.10. Social television response for the *Oscar*s with Topsy.

This number more or less matches the information published on the net: on 27th February 2013, All Twitter (Dugan, 2012) announced that Tweetreach (tweetreach.com, measures how far the tweet travels) measured 2 million tweets just during the *Oscar* night itself.

They tracked Twitter activity on the *Oscar* night and on a movie basis (movies which were nominated for the *Oscars*) analysed the tweets in the Academy Awards explorer (tweetreach.com).

The query for *American Dad* performed on 26th February 2013 produced 49,044 tweets during the previous seven days in total: 10,171 were positive tweets, 7,842 were negative, and 31,031 tweets were neutral. At first, there were no significant results produced by the "*Who Wants to be a Millionaire"* query. Topsy suggested searching related terms such as #millionaire, which was more effective. The title was simply too long, given the 140-character limit for tweets, so therefore only the promoters of this format would use the complete title to inform about the show, e.g. "A rice U student will be on *Who Wants to be a Millionaire* tomorrow!". The impact of this for television program names in the future could be that only short titles or titles easy to abbreviate will be selected.

5.1.5 Social media search with social mention

Social Mention is a social media search platform that aggregates user generated content into a single stream of information. Social Mention (socialmention.com) uses various sources (such as blogs, microblogs, networks, bookmarks, comments, events, images, news, videos, audio, questions) to cover all types of user-generated content. The tool shows the top hashtags (hashtag cloud) and the top key words (figure 5.11), and this reflects whether or not the tool 'understood' the query.

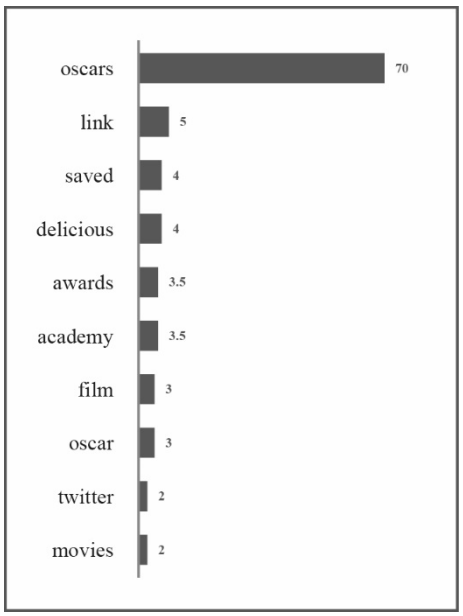


Figure 5.11. Top keyword word cloud for *Oscars*, submitted by Social Mention (2014).

The hashtags shown for the *Oscars* could assist further research. The user would be able to select from a wide range of social media sources. A query with a large number of results delivers approximately 350 mentions, which is more than the previously mentioned tools offered. Social Mention informs the user that the query result is generally limited by the social media source itself. Social Mention informs that "search query data is not currently archived therefore results are limited to the X most recent items from each social media source. Where X is the data limit set by the social media source itself, for example Twitter currently has a data limit of 100 items." (code.google.com, cited in Züll and Mikelić Preradović, 2013, p. 128)

The tool successfully recognised *Who Wants to be a Millionaire* as a full brand and delivered results. However, the search for *American Dad* produced no Twitter or Facebook results, because it did not recognise *American Dad* as a brand. The top hashtags which were listed, such as 'American', that were shown in the results were not related to the series.

5.2 Twitter web archives: Data collection corpus and sampling

This chapter is composed of three parts. The first discusses the approach to a Twitter data collection by possible access to Twitter data collections (library) submitted by Twitter itself.

The second part presents methods where the researcher collects data to build up a web archive based on hashtag research. The data collection can be created by using the streaming API of Twitter which gives the researcher a latency access to a stream of Twitter data and enables all tweets for a certain time period and for predefined criteria to be retrieved. The discussing and testing of the second parts leads to the third one: developing a specific analysis approaching data collections generated through a selective archiving strategy (Lomborg, 2012) to search for and investigate opinions of television audiences.

The best alternative to gaining access to data would be data access (grant) by Twitter. The Twitter data grant is particularly valuable to researchers and without data grants, the access to these data points is hardly possible (Ravindranath, 2014; The Guardian 2015).

The last data grant program was conducted by Twitter in 2014 (appendix 1). The data grant program represents the first formal offer of Twitter to open the vault to researchers, free of charge (Ravindranath, 2014). Within the framework of this research the author applied for access to all Croatian, Serbian and Bosnian tweets for a selected time frame (appendix 2).

Twitter received more than 1000 applications from academic research groups (Ravindranath, 2014). Research applications have been diverse.

Only six proposals worldwide got a positive response: Harvard Medical School/Boston Children's Hospital (US) to study Foodborne Gastrointestinal Illness Surveillance using Twitter Data. The Japanese University (NICT) applied for data to research Disaster and Information Analysis System, researchers at the University of Twente in the Netherlands are assessing the effectiveness of social media campaigns encouraging early cancer detection on Twitter. The University of California in San Diego studies whether happy people are likely post happy images, allowing it to measure the relative happiness of residents. The University of Wollongong in Australia researchers use Geosocial Intelligence to Model Urban Flooding in Jakarta, Indonesia. The University of East London investigated the relationship between tweets and sports team performance. All institutions received free data sets in order to move forward with their research (Twitter, 2014).

The application for data grant was refused by Twitter which leads to the second part of the chapter, the creation of data collections by the researcher.

In general, data collection can be created by using the streaming API of Twitter which gives the researcher a latency access to a stream of Twitter data and enables all tweets for a certain time period and for predefined criteria to be retrieved. The implementation of a streaming client and the standing connection to the public submits a stream of Twitter data which can be analysed by the researcher. The difference in opinion mining tools is the fact that they offer to keep the query and collect and archive any findings over a certain period of time, which has to be defined.

Following Bruns (2011, 2012) the open source software *Yourtwapperkeeper* was selected to demonstrate the characteristics of datasets which could be retrieved by this kind of tool, such as language code or time when the tweet was posted. As the market for this kind of software is in a state of continuous change, one has to expect that this software will not be accessible permanently and therefore simply interpreted as an example for a tool which enables the researcher to retrieve full Twitter data sets.

Once the query is set, the researcher is able to retrieve the data. Furthermore, the data can be exported, for example to Excel, and the researcher is therefore able to filter and work with the data structure. If enough data can be found, automatic text recognition programs can

be used to conduct the next step of the research. The amount of data which can be retrieved by Twitter is currently limited to 1% of all tweets by Twitter. It is possible to analyse various parameters which come along with the export and archive of tweets (Go, Bhayani and Huang, 2009).

However, the interesting aspect of this part of the analysis and in respect to traditional television audience research is the variety of parameters which the Twitter data stream reveals as metadata, as it gives information about users and audiences:

User and user ID includes the profile information of the user who posted the tweet and numerical ID of the tweet recipient (for @replies) (not always set, even for tweets which contain @replies; from_user provides information about the screen name of the tweet sender). The URL of the tweet sender's profile picture submits information about profile pictures of users. Users tweet, create lists and have a home_timeline which can be looked up (Twitter Developers 2015) in detail. This information can be used to research and analyse Twitter accounts. Tweets of television broadcasters could be filtered and identified as they are viewed by the audience for promotional purposes.

Geolocation and iCoordinates represent the geographic location of the tweets (Lettaru, et al., 2013). The iCoordinates of a tweet portray the longitude and latitude of the tweets location, which gives information to the researcher about where the comment towards a television show was posted. Created_at gives information about the time when a tweet was created (UTC time), the so-called tweet timestamp and the time can be used to find out if the tweet was, for example, posted before, during or after the broadcast of a television program. With the language identifier (such as en, de, fr, hr, sl) tweets containing Croatian can be identified, even if the entire tweet is not in Croatian. Text (twitter-text): contents of the tweet itself, containing 140 characters or less and in the text the presence of any relevant television-related information in the text can be identified and analysed.

A potential inaccuracy or bias in the results is that

- Croatian Twitter users may sometimes post tweets in English
- Croat who live abroad may watch Croatian television via the internet and post a comment in other languages.
- Croatian is similar to other Slavic languages: The archive source describes the API source of the tweet, which means Twitter–search or Twitter–stream.

The first queries were created in November 2013. Again the examples *The Oscars, Who wants to be a Millionaire* and *American Dad* were selected for the queries. The number of tweets for *Who wants to be a Millionaire* and *American Dad* was below ten and not taken into consideration for further analysis. The analysis for *the Oscars* will be portrayed in detail:

The following search terms were set (table 5.2) as for the research of *the Oscars* and it was defined that several hashtags/keywords would be selected (learned from working with Topsy) as a keyword cloud, to retrieve more data:

Table 5.2. Definition of search terms.

Hashtag	Description
#redcarpet	Oscars red carpet
#RTLOskar	RTL Oskar
#Oscar	Global as the term is written with "c"
#Oskar	Oskar HR, written with "k"
#TheAcademy	Other description for Oscars

The query for these search terms generated 729 tweets. Within the four weeks around the broadcast date of the Oscars in 2013 there were 729 tweets retrieved.

By examining the search tweets containing *#Oskars*, it was discovered that none of these Tweets was marked with the Croatian language code and that it was not possible to assign the results of research to the broadcast of the television show at HRT in Croatia. Different languages were found, which just confirmed the fact that the show was broadcast all over the world. Surprisingly, not a single tweet was marked with hr for Croatia. The biggest share of tweets (figure) was created in Italian (303 tweets), followed by Russian (80) and English (87 tweets).

Table 5.3. Distribution of tweets following language code (University Penn State, 2015).

Language	Tweets	Language	Tweets
Arabic (ar)	2	Italian (it)	303
Bulgarian (bg)	13	Latvian (lv)	46
German (de)	60	Norwegian (no)	1
Greek (el)	2	Polish (pl)	17
English (en)	87	Russian (ru)	80
Spanish (es)	7	Slovak (sk)	2
Estonian (et)	2	Slovenian (sl)	77
French (fr)	1	Swedish (sv)	9
Former Indonesian (in)	6	Turkish (tr)	3
Islandic (is)	1	Undetermined (und)	10

As no Croatian tweet could be found for *the Oscars*, the analysis was also stopped at this point for *Who Wants to be a Millionaire* and *American Dad*.

Open-source platforms utilise the Twitter stream, in almost real time, and capture tweets containing selected hashtags. When the broadcaster introduces hashtags, the discussion on Twitter could be more easily located. A so-called word cloud – usually a group of hashtags/keywords– exists around a television program, such as the title of the program, its main cast, or episode titles.

If television broadcasters no not promote the communication about television programs and the television audience does not use hashtags/keywords to mark the discussion, then only weak results can be expected.

The third part of this chapter presents data corpus work generated through Ljubešić, Fišer and Erjavec who used an open-source tool called TweetCaT which was designed to collect tweets and build corpora for (smaller) languages.

The tool used *language* as an identifier for Croatian, Serbian and Slovene tweets. Building and/or using an 'existing' data corpora of Twitter data is not a new phenomenon: Ljubešić, Fišer and Erjavec (2014, p.1) refer to the Edinburgh Twitter corpus that contains almost 100 million tweets, and the Stanford Twitter Corpus with 467 million tweets. The Edinburgh Twitter corpus was used by Zhao, et al. (2011) to discover media related topics and for trend detection (Benhardus and Kalita, 2013) from a representative sample of the entire Twitter (data collection).

The data sets of Twitter include metadata which means additional information parameters accompanying the text of the tweets which were considered to be relevant to the analysis. Two corpora were build, one for Croatian and Serbian and one for Slovene using "seed terms and simple language identification module to find new users as well as new tweets from already known users that tweet in the target language" (2014, p. 1).

The use of a tool such as TweetCat enables the researcher to collect tweets in any language through Twitter API and is therefore suitable for building Twitter data corpora based on specific Twitter characteristics (p.1). The corpus did not discriminate at the beginning between Croatian and Serbian because of the complexity of the task (p. 1). The data was collected over a period of 235 days from 1 June, 2013 to 21 December, 2013. The tool identified and collected tweets and produced a "constant and significant stream of data" (p. 2) during the time period. The result of the data collection was a data corpus of 26 million tweets. For a more detailed explanation and more information about the applied method which was used to build the corpus see Ljubešić, Fišer and Erjavec (2014).

The idea of sampling is well known from traditional audience analysis Zhao, et al. (2011) to get a representative sample for the data corpus. Once the researcher has access to a data collection, one could discuss the representativeness of Twitter data in respect to all television audiences which is not given.

In an ideal world the data corpus would be freely accessible by researchers and could be considered a public library, as has already been mentioned. The access to data would not be controlled by data 'gate keepers', such as Twitter who cap the supply (sample of public stream) at 1% of the stream. However, in this research a data corpus of 26 million tweets could be used for analysis, which was considered a significant size for analysis. Methods such as TF-IDF based bag-of-words can give a term-level topic modelling (Michelson and Macskassy, 2010), which is appropriate for ranking media related key words and therefore for discovering the comments of interest.

A text editor software tool (sweetscape) was used, which enables the researcher to export and search big data. The tool enables the researcher to analyse and edit binary files (which are readable by a computer). The number and size of samples was determined randomly, and it would be necessary to investigate the size of the data corpus and the number and size of samples in further research. It was therefore decided that 10 samples to locate media and television-related tweets would be taken, with each sample containing 10,000 tweets. In total,

100,000 tweets were analysed with a model developed step by step during the empirical research to rank media related terms based on term frequency, similar to news (content) ranked articles on Twitter (Phelan, McCarthy and Smyth, 2009).

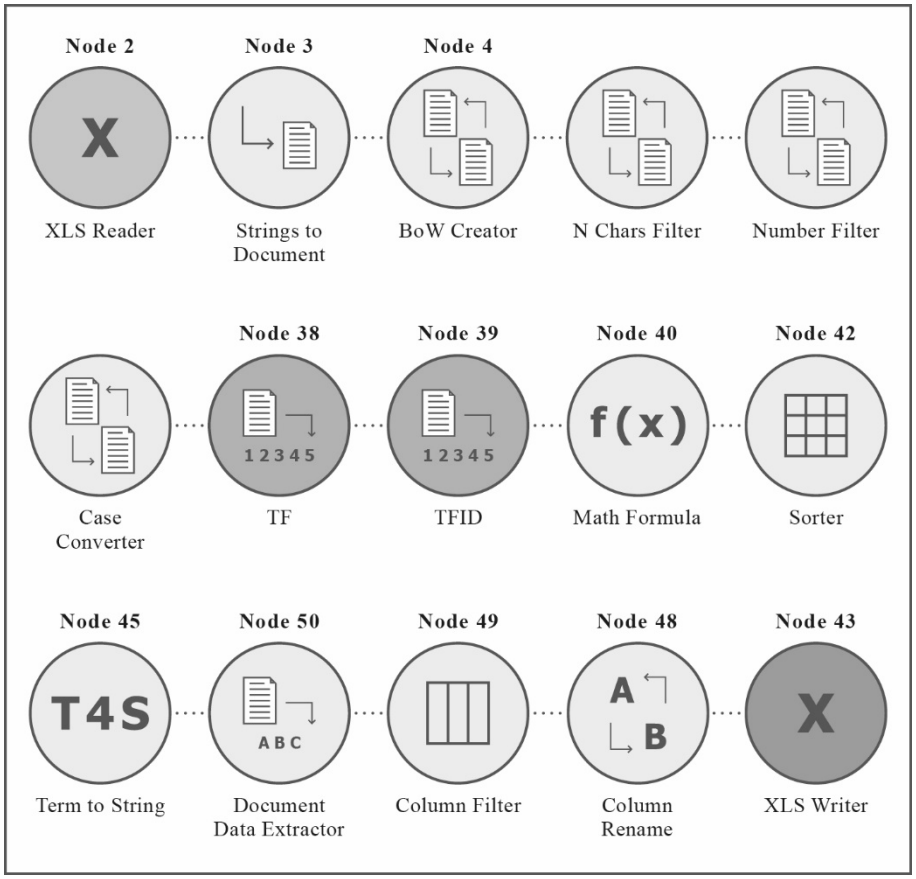


Figure 5.12. Location of media and television-related tweets in a data corpus (Michelson and Macskassy, 2010; Knime Analytics Platform, 2014).

The location of media related content was performed with a XLS reader, seen as node 2 in figure 5.12, a simple freeware for viewing spreadsheet documents created by Microsoft Excel.

To identify and look at each word in the tweet, the words of a tweet need to be separated. The separation of the words of a tweet can be done applying a "bag-of-words" model (node 4) which is a simplifying representation used in natural language processing and information retrieval (Michelson and Macskassy, 2010). The text, in this case the tweet, is represented as the bag of its word, disregarding grammar or word order but keeping multiplicity. The tweets are divided into single words with a bag-of world creator. The bag-of-word model is commonly used in methods of document classification, where the frequency of (in this case television-related) words is used as a feature for training a classifier (Zellig, 1954). Although

Billsus and Pazzani argue that content-based approaches to finding content on Twitter are difficult because of the sheer random bag-of-words unstructured nature of tweets, and the complexity of natural language processing (1999).

To clean the tweets all words with less than 3 characters a chars filter was applied which means all words were deleted with less than 3 characters (KNIME, 2014; Bakos, 2013; Thompson, Mo, Pacheco, and Carroll, no date), with the purpose of keeping only words with at least 3 letters. Words relevant in the context of television were manually checked, and a cleaned version was prepared. The difficulty is to manually collect enough data to train a sentiment classifier for tweets (Go, Bhayani and Huang, 2009).

All numbers were removed and all words in uppercase were converted to lower case (number filter and case converter). Finally, the term frequency-inverse frequency (tf-idf value) that reflects how important a word in a document is to a query (Benhardus and Kalita, 2013), in this case a tweet (node 38 and 39) was applied within the data collection.

The tf-idf value increases proportionally to the number of times a word appears in the document (Benhardus and Kalita, 2013). It is a simple way to determine which tweet is the most relevant to a query. To make a further distinction, one might count the number of times each term occurs in each tweet and add them together. At the end, the text (in this case the media and television related keywords) can be read and extracted by an xls-writer (Bakos, 2013) and a ranking list (table 5.4) is submitted based on the samples retrieved from the data collection.

Table 5.4. Top 30 - Tf-idf ranking of television-related tweets.

rts2	1,34
farma	1,34
#bigbangtheory	1,34
#despeterhousewifes	1,34
tv-u“	1,34
ezele	1,34
#b92	1,18
hrt3	1,18
.@rts_vesti	1,03
ibrahime	0,89
watching	1,03
foxlife	0,89
.@tv_arenasport	0,79
sulejman	0,79
.@rts_vesti	0,69
policijska akademija	0,89
#eurotrip	0,67
dnevnik	0,67
animal planet	0,67
exkluziv	0,67
episode	0,67
poison ivy	0,67
crtace	0,67
greyanatomy	0,67
film	0,53
nu2	0,53

This ranking list submits a group of television related keywords which can be used as the basis for further research as these keywords represent the most significant in the data corpus. The first two positions are the most used words in tweets related to Serbian media.

The following international television shows could be discovered as a result of this research:

#Eurovision which is a European song contest broadcast every year in all European countries; the *American* sitcom #bigbangtheory, which is broadcast on RTL and RTL 2 in Croatia, B92 in Serbia, Nova Television in Bulgaria, Slovenia on Kanal A, and many other countries in Europe; and #Desperatehousewifes, an American television series which was broadcast all over Europe, such as Grey’s Anatomy.

The following television channels could be identified: HRT2, RTS2 (Serbian television channel), #b92 (Serbian television channel), HRT3, and Fox life.

For local television programs, only the Croatian television show *Farma* broadcast on Nova TV which debuted in 2008 could be identified at position number six. For television series, the following titles could be identified: *Sulejman* stands for *Sulejman the Magnificent,* one of the most successful television series in Croatia and ranked among the top 10 television series in Croatia*;* and *Ezele* stands for the Turkish drama series *Ezel*, broadcast in Croatia on RTL Hrvatska. As *Sulejman the Magnificent* ranked among the top series it was decided to look into this television program in more detail and the full data corpus was used again with the search key word #*Sulejman near: "Croatia"* (description: Sulejman blizu HR). The hashtag was amended with "near Croatia", to ensure that only tweets in respect to Croatia would be collected. The search in the database of Ljubešić, Fišer and Erjavec (2014) submitted in total 1,867 tweets which contained the keyword Sulejman.

It was discovered that most tweets for *Sulejman the Magnificent w*ere created by the television channel RTL Televizija, with the purpose of promoting the series. The tweets contain links with information about the series or announcements of the episodes. It was considered to be a surprise that in the metadata most tweets contained the language code sl, which stands for Slovenia, although the author of the tweets was always RTL Televizija. It can be assumed that as Slavic languages are similar it is difficult to assign the languages to the individual countries.

Furthermore, some of the authors were from Serbia, not from Croatia. The reason is that *Suleiman the Magnificent* was broadcasted in Serbia as well. This shows how difficult it is to assign social media data to the television viewers of the broadcasters, in particular when the series is broadcasted parallel in neighbouring countries. All tweets in the extract were created during weekdays, and the majority of tweets were created during daytime or close to broadcasting time (prime time). It was only observed once that geo information was attached to the tweet. It is possible to search for users, to continue the analysis by searching on Twitter and to find a place of residence in their profile information.

Only a minority of tweets were posted in the context of viewing the series, therefore the research produced only minor insights into the communication activities of the television audience before, during or after the show. Most of the tweets were posted by public media institutions for promotion purposes.

This analysis does not lay any claim to completeness but only demonstrates a different way of approaching the data. Some television-related key words could be found during the process of sample-feature representation.

6 EVALUATION AND MODELL

This research started with the depiction of the television landscape in Croatia, focusing on the transformation of television broadcasters, audience and user habits primarily caused by the development of new technology and the internet. With steadily increasing development of television media and early digitisation over recent years, Croatia is a reference market for all of Southeast Europe (chapter 2).

Social media is characterised by disclosure (Haenlein and Kaplan, 2010; chapter 2.2) and the internet users in Croatia and worldwide (chapter 2.2.1) upload pictures and videos, post tweets often enriched with emoticons and messages, use social networks and search engines, leave an enormous amount of personal information and private data on the internet and affordable data storage, and their global reach creates a global memory which enables retrieval of user data. Content communities and social networks ask the user to disclose information about themselves as a result of the sharing of content.

Television audiences use social media to participate in television programs often described under the term social television. Users communicate about a television program, they share information and opinions and by sharing content they submit – consciously or subconsciously – metadata which reveals details (5.2) which can be analysed by researchers. Users are aware that they leave traces on the internet, but probably have little or no knowledge about how their data can be used for research, with both commercial and academic interests. With the introduction of "digital communication systems, a new type and scale of data are there to be found: big data or metadata that indicate who did what, with which information, together with whom, when, for how long and in which sequences and networks" (Jensen, 2014, p. 229). Once the content is uploaded and shared, it can be found and used for research (Züll, 2013).

Expansion of social media activities influences all areas of life in particular for younger television audiences in Croatia (chapter 3). Users spend time on social media and television broadcasters as huge content provider submit topics to share on social media platforms. Public and commercial television in Croatia transport content to their audiences depending on their (public) mandate e.g. to inform, educate and entertain and it has always been a concern of the broadcasters to learn if the audience as a receiver of the broadcast appreciates their program. Television audience research is justified by the need to know if television viewers watch and like the program. The television audience measurement based on television

ratings has been the main currency of the industry for decades (chapter 3) and although television markets are in the process of fragmentation and user habits (chapter 2), especially of young audiences, are changing, there is no doubt about the importance of this tool. Traditional audience measurement submits daily, on a minute by minute basis information for every program. Although often criticised, the system which generates this data is sophisticated and accepted by all participating parties. In particular, the statistically demographic relevance of data in combination with qualitative information about panel members (investigated in chapter 3) submits a pretty complete picture of the television audience. Despite devices employed for audience measurement during the last decades, it seems that it is not enough for media practioners, for market researchers (Vincente-Marino, 2014) and academics (e.g., Cost Action ISO906, Transforming Audiences, Transforming Societies), who search for deeper information about television audiences. The measurement system provides information about viewing patterns and demographic composition of television audiences. Morley examines four main assumptions of television ratings measurement (Morley, 2010) and stresses the need for deeper multidimensional research of television audience behaviour:

- 'Reliable indicator viewing' is if the television set is on or off and the viewer is presenting the room, but there is no information about whether or not their attention is on the program/channel.
- Switching the television on is equalised with the conscious selection of a certain program/channel, although the decision can be a 'reflex-reaction' signifying 'getting home'.
- The assumption that the viewing behaviour is the result of individual decision: group decisions are not taken into consideration, although the decision making is often more dominated by some family members than by others.
- The fact that television ratings measurement ignores the significance of viewers' contextual factors such as variation in the amount of household space, and income level.

Morley (2010, p. 168) cites Jensen (1987) who furthermore stressed the need for "contextualisation of research 'findings'" and argued that "[the ratings] offer view clues for

understanding the significance of television as an integrated element in the viewer's everyday life. The audience experience of a particular medium and its content cannot be separated from how it is used […] if we are to understand the lived reality behind the ratings we need to turn to the context of use, the physical setting where reception takes place, what is the meaning of TV viewing to the audience."

Hence, social media opens new ways for research methods to approach data, particular opinions and attitudes of television viewers are identified within the frame of this work to be a gain in addition to traditional research approaches.

Opinion towards a television program can be mined, although one does have to organise the access to data, which is apparently difficult. One can carefully make the assumption that even if researchers pay for social media analysis tools to get access to data, it remains difficult to interpret the results as algorithms and retrieval of data remain a challenge.

In general, Twitter proved to be a suitable social media network and data source for the research, even though the usage numbers of Twitter in Croatia are smaller compared to Facebook, for example, which is a common phenomenon all over the world (chapter 2.2.3). The researcher cannot control the Twitter API and a conclusion with certainty regarding the composition of data cannot be made. Although it was discovered that there is a lack of transparency, the Twitter API was used within its limits (compare Bruns & Liang 2012). Compared to the United States or other European countries, social television strategies are still underdeveloped in Croatia.

Only a minority of the audiences in Croatia uses Twitter. However, one could argue that younger people particularly use social media platforms, and therefore the results are nevertheless interesting to the industry.

The general advantages for the researcher analysing social media data is clear: social media data, in particular opinions and comments which can be found on the internet are real time and the researcher can access personal statements which are not filtered or influenced by an artificial research situation, such as interviews.

Furthermore, social media data is archived and can be researched weeks, months or years later. The approach to analysing social media data could be a possible way to prove if broadcasters reach younger audiences and will determine whether it will be possible to keep their relationship to younger audiences.

The empirical part develops and applies different methods to approach Twitter data and search for audience opinions towards selected television programs, to find out if social media analytics in the context of television consumption can be a new valid source of information in Croatia (chapter 5).

The study was divided in two sections, to create a methodological framework which would have the potential to validate and find answers to the hypothesis of this work (chapter 4.3).

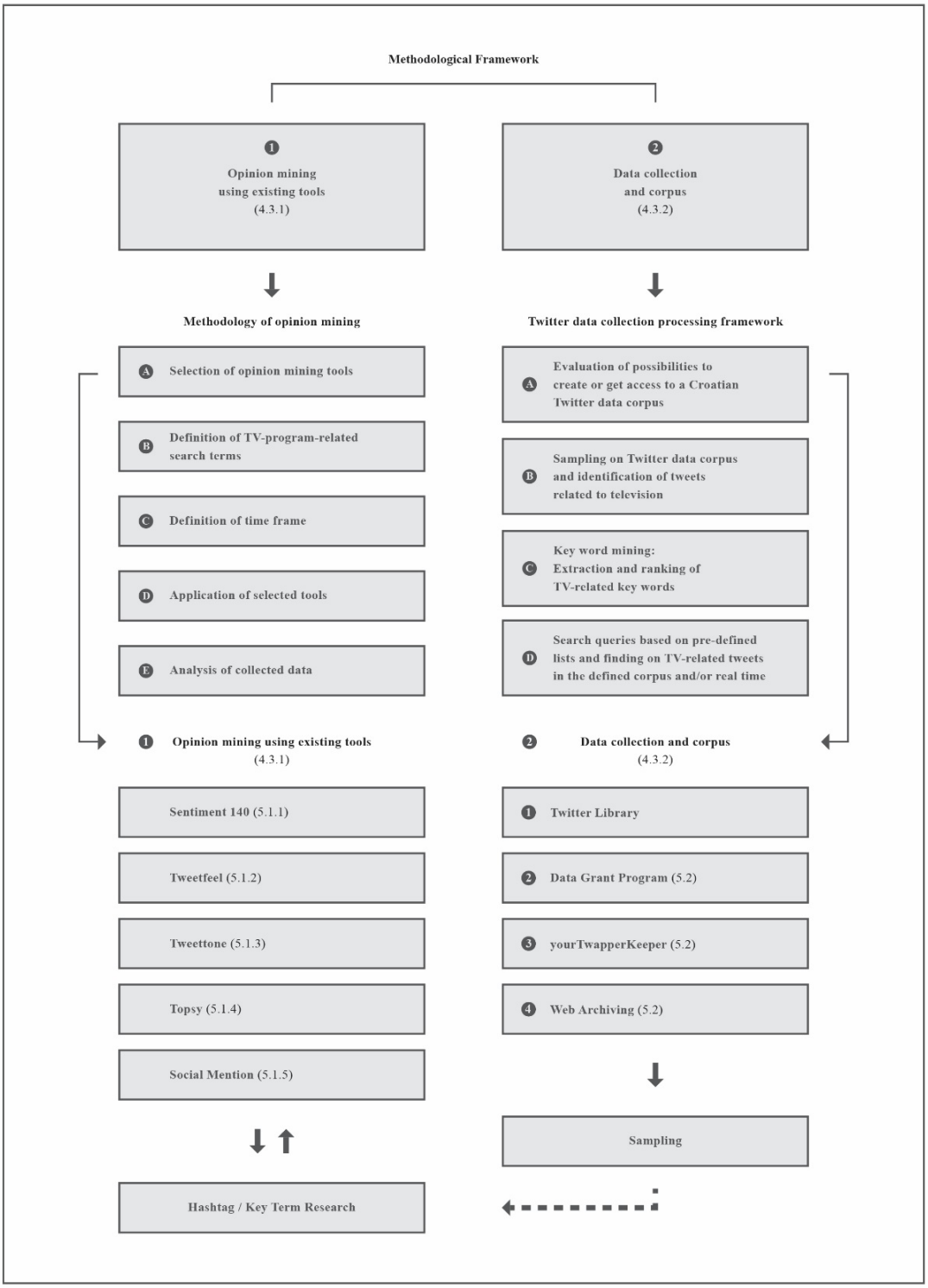


Figure 6.1. Methodological Framework.

The method of working with opinion mining tools was used to mine opinions of Croatian television audiences with key word/hashtag searches.

Twitter data collections were considered as useful, secure instruments for producing more confidence in the interpretation of data and sampling was introduced as a possible approach to finding media-related key words in huge data collections.

The initiative of the Library of Congress and Twitter was considered as a pioneering project. Twitter as new source of a kind of digital data collection and library should – following the Library of Congress (appendix 3) – have the mission of serving the Croatian public. As Croatian society turns to social media and uses social platforms as the primary method of communication, sharing and creative expression, social media is supplementing sources routinely collected by research libraries. The data grant offered by Twitter 'Intoducing Twitter Data Grants' (appendix 1) paved the way for the idea of getting data by Twitter for Croatia, and therefore it was applied within this research by the author (appendix 2).

However the central thread woven throughout this work is the access to Croatian social media data to study television audience behaviour.

Communication on Twitter needs to reach a critical mass to be found and analysed and even to be used to discover and interpret patterns of social media usage. Only strong television brands therefore seem to show a social response on Twitter. Different television programs evoke different communication activities (Buschow, Schneider and Ueberheide, 2014). The quality of results can only be as good as the search query within a selected timeframe, as certain keywords may not appear during the Twitter research. Television broadcasters need to promote hashtags or selected key words of the program on air to initiate online discussion among audiences. Otherwise discussion cannot be located by users and researchers.

The empirical analysis of Croatian data submitted only limited results, however the question of how to approach social media data was considered as an even more important question during this work.

The first hypothesis of this research that social media plays an increasingly important role for the Croatian television audience, giving people the opportunity to share their viewing experiences in various forms and on various social media platforms could be confirmed under the assumption that as people in Croatia use social media and social media usage numbers steadily grow, the opportunity to share and discuss television content is automatically given.

The second hypothesis is that a relationship can be found between television consumption and social media activity could not be confirmed for Croatia as not enough data could be found, and furthermore it was not possible to assign the data to Croatian television channels. Only tweets which were posted for promotion purposes could be found.

The third hypothesis, that social media activity can impact on television ratings, and therefore influences the success of a television program is a phenomenon which is discussed in an international context and some references and results of other research groups could be found (chapter 4). This does not apply to Croatia as the Croatian television broadcasters are still in the beginning of the creation of a social television audience. Croats use social media to communicate, it is not proven yet that Croats follow the social media promotion activities by the broadcasters. However, initiatives by broadcasters to promote and stimulate communication among audiences regarding television programs is needed.

The analysis of social media activities in the context of television seems conditional upon online activities of television broadcasters and if the Croatian television audience follows them, to maintain and to build up relationships with younger audiences via social media, similar to the BBC in the United Kingdom (Department for Culture, Media and Sport, 2006). Therefore the second part of this chapter is as conclusion dedicated to the question of how a social media audience could be built up by Croatian television broadcasters as a precondition for the analysis of social media activities which are related to the television program. Social television analytics measures not only sentiment, listing a percentage of positive and negative comments, but volume, measuring the number of comments made in social media.

6.1 Tracking popularity with social media analytic tools

Television content is dynamic and public sentiment results may change quickly, depending on the time slot, programs broadcasted on television and also the moment of analysis. The result of Twitter opinion mining represents a snapshot of a sentiment at a certain moment in time.

A selected number of opinion mining tools and possibilities to retrieve data were applied. It is not a surprise that varied sentiment results were produced.

The issue of finding opinion sources and monitoring them on social media is a challenging natural language task due to a large number of diverse sources. The main problem

of automated sentiment discovery is to find relevant sources, extract key information and opinions and assign them to a selected television program.

It should reveal the attitude, expressed opinion, emotions or evaluation of the audience and is used to classify the comment to determine whether they have positive or negative polarity. Audience research has fuelled interest in opinion mining, to analyse the web communication. Personal opinions in the form of reviews, ratings and recommendations and other forms of online expression deliver information about television viewers using Twitter.

A disadvantage of statistical-based approaches to social media data is failing to integrate subtleties such as irony or sarcasm. On the other hand, linguistic approaches use libraries containing words and extract the semantic information that represent attitudes to indicate an opinion about a television program.

The challenge for opinion mining tools is to develop sophisticated algorithms and software that look at several different filters, including polarity, intensity and irony. The quality of tools depends heavily on the openness of the provider in respect to methodology and transparency of the data source. Additionally, it is important for the researcher to have information about the data volume which is processed by the provider. But which tool delivers accurate results and how can an algorithm used by a tool be evaluated by a researcher or user? How are different lexical indicators of sentiments in opinion sentences evaluated? It was discovered that some opinion mining tools process different volumes of results, either depending on their restrictions set by Twitter, their own data policy or the algorithm they use. Does this mean that Twitter sets different data limits for different tools and suppliers?

These questions are challenging to answer and depend on the transparency of tool suppliers in respect to the methodology used and information about data sources. It can be concluded that the data source has a huge impact on results. It is necessary to perform further research about the limitations and origins of data (territory, information about time and rolling basis of data, volumes of data).

There is therefore no definite answer to the question “which tool performed the best”, because the evaluation depends on various criteria: the source of the data used, algorithms, the number of results, territory, number of possible keywords, filter options and timeline.

Only when tools produce relevant hashtags, further research can bring additional results. The analysis showed that not only the keyword connected to brands or the title of the

show is significant and that the choice of keywords for queries is of high significance. It has remained a challenge for researchers and users to define adequate queries to find and identify the discussion related to a certain program. How can the most appropriate Twitter acronyms related to a television program be found? Is possible and necessary to link video streams to the social media comments? How can users locate content related to a certain program?

One can choose the title or the keywords introduced by the broadcasters but how can it be ensured that these keywords are adopted by the Twitter audience to talk about the specific program? How can the researcher then be sure that the selected keyword search matches the majority of comments and discussion? What can be done if the broadcaster does not use hashtags to promote its programs? How can the opinion changes on Twitter be tracked with the opinion mining tools? How can a query be created if the keywords are not known in advance?

Retrieving datasets of users down to the details such as IP address, user profile and location blurs the privacy line and limit of the audience. "Users can be identified via their internet service providers (IP-address and log files), and cookies, search engines, electronic commerce, HHTP protocols and, browsers all track user data. In order to protect privacy, companies often make this data anonymous but studies have revealed that they can still connect it to an individual" (Mendel et al., cited in Züll, 2013, p. 124).

Once the audience is stimulated by television broadcasters (chapter 6.3) it could be possible that the data could be used and analysed for research purposes, and that the data could be attractive to the industry as it seems to be the case in Australia, Italy, Mexico and the United States. The activity of the audience on the internet in the context of their television viewing in a changing and transforming media environment became interesting to researchers, however the "glass human being" and glass audience is no longer just a literary metaphor, but has turned to be a real prospect which raises questions which should be discussed in a much broader scope.

In countries where Twitter usage is just beginning, usually the usage (by users and broadcasters) of hashtags has not been established yet. Television channels in Croatia did not introduce and communicate Twitter search terms to their audience until 2014 (Züll, Boras, Mikelić Preradović, 2013). The key goal for broadcasters is to ensure that the audience automatically finds the program they share. If broadcasters do not develop social media activities to promote the communication around their programs, it remains difficult to build a social media audience. The queries were performed several times on a selected day and were changing during the course of the day. In addition, numbers of tweets seem to increase during

the course of a day, and therefore a correlation between the social media activity and television consumption could be theoretically assumed.

The tools teach the user which queries will deliver results and which will not. Furthermore, long titles such as *Who Wants to be a Millionaire* cannot deliver a sufficient number of results.

Only if tools show relevant hashtags, further research can bring additional results. Additionally, it means that defining appropriate queries represents a challenge. What is the best way to find the query terms used by the television audience discussing in social media? How to find the best Twitter acronyms related to a television program? Is there a need and possibility to tie video streams to social media comments? How can the user find the discussion related to a certain program? How to track the opinion changes on Twitter with the opinion mining tools?

In this testing of opinion mining tools only foreign comments about the *Oscars* could be found so no relationship between tweets and television consumption for Croatia could be discovered.

In the case of the research performed for the *Oscars,* the discussion and comments which could be found were related to a variety of topics such as the movies, the music, the outfits and performances, predictions about the winners, and so on.

The Oscars

The Oscars turned out to be the 'best' example, as a huge social media response was revealed in the analysis. It can therefore be assumed that a significant amount of data is needed in order to find reasonable information about television viewers' attitudes towards a show. The queries for the *Oscars* submitted reliable results as they matched up to the official numbers published by other sources. The query using Topsy produced 4.3 million tweets. The Oscars had patterns around the broadcast time of the show, which were similar to the patterns revealed by television ratings analysis.

Who wants to be a Millionaire

At first, the query for *Who Wants to be a Millionaire* produced no significant results, but by using Topsy the top hashtag #millionaire brought more results. This reveals that the

title of a television program in the context of social media is important as it needs to fit in a 140-character post, and should therefore be short and easy to remember.

American Dad

The challenge in analysing *American Dad* and *Who Wants to be a Millionaire* was whether or not the tools could recognise the name of the show as one brand and have the ability of finding connected key words. The *American Dad* query was successful and produced a large amount of information about television viewers and their social media activity. How could the selection process be improved?

Topsy promotes its tools for researching television networks and therefore seems to be – as a commercial tool –very attractive for television media researchers. What happens if users introduce other key words or abbreviations to organise their communication and change some of them over the time? Some tools, such as Social Mention, show as results of the query a hashtag or keyword cloud, which enables the researcher to locate more communication regarding a television show.

The analysis demonstrated how difficult it is to assign social media data to its own social media audiences. How can it be ensured that the comments of the social media audience can be assigned to a television channel? This is particularly problematic when the television program is shown simultaneously in several counties (worldwide television events), in the same language.

The data which can be captured using a tool such as *Yourtwapperkeeper* contains the text and other essential information about a tweet – this is the metadata of the tweet and contains a number of elements. In particular the identification of the location where the tweet was posted (in cases where users allow the geo-location), the language and the geotagging can provide information which connects the tweet to a television broadcaster and its programs. The tweet itself is of interest particularly when the language gives an indication of whether the tweet was created in the television market of interest.

In particular, the origin (where the tweet was posted) characterised by language code and geotagging could be seen as a characteristic by which to assign the tweet. None of the tweets were written in Croatian, none of these tweets were marked by a Croatian language code or by a corresponding geolocation. The conclusion therefore is that either no television related

discussion can be found, or that Croatians do not use hashtags or particular keywords to mark their discussion.

6.2 Opinion mining working with data collections

For the social media analysis it is important to get more information and as much data as possible. As the usage of opinion mining tools based on hashtag and keyword search submitted only limited results for Croatia, it was decided to approach the analysis of data by working with a bigger data collection, by attempting to gain access to a bigger collection of tweets for Croatia. In general, the access to a Croatian Twitter library would enable Croatian scholarship to access a fuller picture of cultural norms, dialogue, trends and events as is demonstrated by Twitter and the Library of Congress in the United States (2013). Although the main challenge for librarians would still be the development of technologically sustainable infrastructure for the digital data collection and archiving task. No other country worldwide has tried to deal with that digital challenge yet and there is a need to observe how and when this task can be managed in the U.S. library. At first there was an application gain access to data via a Twitter data grant program in 2014, but the University of Zagreb was not taken into consideration (5.3; appendix 2).

Ljubešić, N., Fišer D. and Erjavec, T. (2014) published their development of a tool for building Twitter data collections (Twitter corpora) for smaller languages which was used for this research. The idea to approach the data corpus via sampling was developed, as it was revealed that the opinion mining process would need to find and consider the word cloud of queries to achieve more accurate results, but at the same time needed to assume that the word cloud was by no means complete. Furthermore, the following question arises: how can the Croatian television audience participating in social media organise their communication if Croatian broadcasters do not publish keywords or hashtags and as part of a social media campaign around their brands and television programs?

Ljubešić, Fišer and Erjavec (2014) built a data collection (corpus) for Croatian and Serbian tweets which was used for this research. It was interpreted as an advantage for this research that the tool enables the researcher to access a collection of Croatian tweets – including metadata – which was collected over a time frame of 235 days and which produced a constant and significant stream of data, although being aware that that the collection itself is due to the Twitter restriction to provide only 1% of all tweets though the API sample.

To test the approach a number of samples was determined randomly and 100,000 tweets of the collection were analysed. It would be part of further research to statistically determine the optimal size for a defined Twitter data collection. Although the software used was still considered automatised standard solutions, the research came to a point where words and terms relevant to media in the context of television had to be checked manually.

One can hardly imagine that the measurement of social media activity could be performed manually, in particular in growing and more and more fragmented markets. However, the result of this research was a ranking list of keywords which could be used for further research, and by comparing the results of the analysis of Twitter data (ranking list) with the top 50 list of television programs in Croatia (table 3.2), the Turkish television program *Sulejman* was discovered on both lists. Therefore, this title was used as a keyword to search the database and the result of data retrieval showed that most of tweets were published by television broadcasters to promote their program (and not by users). This analysis does not lay any claim to completeness but only demonstrates different approaches to the data. As television related keywords could be found through the process of sampling, it was considered to be a possible approach to finding relevant keywords for analysis which could be used to search communication on Twitter.

In countries where Twitter is established as a backchannel for television, the titles or abbreviations of titles are usually promoted by the television broadcaster as they have an interest in creating conversation around their shows as a part of the viral marketing strategy.

As a conclusion, the following chapter describes the possible social media activities and engagement of television broadcasters because the "revolutionary new trend should be of interest to companies operating in online space" (Kaplan and Haenlein, 2010). However, the organisational-level research on social media activities has not grown as rapidly (Lovejoy and Saxton, 2012).

6.3 Development of social media activities of television

The television industry in general produces and distributes television content in different formats and for different channels and is practically predestined to produce content which can be discussed and shared on social networks. Traditional mass media is the number one topic generator, and audiences talk about programs which they have seen on television.

Media producers and consumers are participants in a multimodal, multichannel system of digital communication that integrates all forms of media (Castells, 2009).

Digital media makes information accessible to audiences and allows people to become senders (Jenkins, 2006). Participants such as television broadcasters remain power communicators, using all channels to distribute their content. Furthermore, individuals within corporate media such as hosts and other selected members of television news rooms interact with audiences.

When the data for the empirical part of this work was collected, the analysis showed that none of the researched television programs had a social media strategy at the time of investigation and that Croatian broadcasters did not stimulate online discussion via, e.g. referring to and broadcasting Twitter hashtags within the broadcast.

Furthermore, the result of the empirical work leads to the assumptions that users of Croatian television had just started to develop towards a social audience and does not follow certain rules or patterns to organise their communication on social media platforms.

In general social media platforms seem to be the ideal partner of broadcasters and for the development of social television as they support the digital social infrastructure which television needs to communicate with its audiences. The participation of television broadcasters and its viewers in social media has changed their relationship (Levine, 2013). Television functions as "society's mirror, reflecting its trends and also, in a kind of strange synergy, shaping them" (Danesi, 2002, p. 139).

The measuring of social responses towards a broadcast raises the question about the role of the broadcaster as creator of the social relationship to particular younger audiences. Television becomes social where television broadcasters create shows and build relationships with the viewers by providing (social) interaction around its program and where millions of people share their television viewing experience.

Most of the television production companies start to have a two pronged approach, where mainstream television programs are increased in social value by interactive elements and furthermore producers start to integrate and test new interactive elements as new concept for television programs (Buschow, Schneider and Ueberheide, 2014).

Thanks to the increasing opportunities for audience participation, the media changes their production and distribution routines, they react to this process and audience participation

is becoming a way to improve media products (Vivo et al., 2013; Andrejevic, 2008). Comparing and contrasting the community of television broadcast and community of internet interactivity reveals that there is no clear distinction between these two (Holmes, 1997; Green, 2001). Therefore, the basic difference between transmission of audio-visual communication through television and through internet lies in the fact that the latter can act as a platform for the production and processing of information and has the potential of giving the user a voice, allowing them to upload text and video messages (Kim and Sawhney, 2002).

"Social media introduce substantial and pervasive changes to communication between organizations, communities, and individuals which presents an enormous challenge for firms" (Kietzmann, et al., 2011, p. 250).

Broadcasters are challenged to establish a new presence in social media with Facebook pages, Twitter accounts and YouTube channels and designed new forms of television-related content such as pictures, texts (editorial or user generated), videos (editorial or user generated, short or long, music video or other news or videos related to a television program) and links. "Integration is key! […] It's all about participation, sharing, and collaboration, rather than straightforward advertising and selling" (Kaplan and Haenlein, 2010, p. 65). It is necessary to integrate television and social media activities (Kaplan and Haenlein, 2010) as television is the most effective tool to reach the audience and to motivate them to start social media activity. Twitter and Facebook all seem to represent different and separate online worlds. However, it can be assumed that there is actually an overlap of social media users who all make up one television fan community which follows its favourite program over all platforms.

The implementation of change management would be a precondition if the potential of social media is to be exploited (Kreutzer, 2104). Kreutzer describes 4 steps of development for companies who decided to use social media as a tool to build up a social television audience. It starts with an experimental phases (step 1) where social media is only partially and randomly used, and where no budget is allocated to social media activities and towards the building of social media structures which get more and more professionalised and established (step 2 and 3), to integrate social media activities within the whole organisational structure of a company (step 4).

Television channels may use social media as (multi-) distribution channels for their work. They provide a variety of information such as program information, breaking news, facts, pictures, reviews and competitions. They allow their audience to participate and share, to

immerse themselves into the programs and to express their opinions and emotions. The broadcasters use social media for news gathering and for building manifold numbers of social media profiles to communicate with the audience and users.

News tends to be retweeted including links to the relevant articles (Boyd et al., 2010).

The official social media (Twitter) accounts can be understood as a new form of many-to-many communication between television and audience. The accounts can be used as a way of content distribution by the television channels and at the same time as a back channel for television: for example, Twitter news may automatically be used as a source of journalistic information, which in some cases has been met with criticism from the public. The usage of Twitter as a source for journalists, networking with audiences and politics to enhance their information sources resulted in heated public debate about the quality of journalism in Croatia and elsewhere. Social media gives the journalists fast access to international news and therefore could also be a chance to establish higher standards for the quality of Croatian journalism, which is according to Malović (2007) still endangered. "The social media guidelines are intended to ensure, that stories are fed into the television channels' news gathering machine as quickly as possible and without the delay of a 140-character update on Twitter" (Plunkett, 2012).

As the audience needs to be guided in using and applying social media, for example using promoted and broadcasted hashtags which mark social media activity and conversation around certain television shows, it will be interesting to observe and further research if the Croatian audience adapt to the hashtag system which would be a precondition for television audience researchers in Croatia to locating relevant data (chapter 5). The BBC, for example, is working on a prototype implementation which will "use the Universal Control API to extract information about what programme a user is watching on their connected television. The extracted information is then used to automatically 'pull' programme relevant content to a companion device such as a mobile phone, tablet or laptop. The user can then choose to share the online content on social networks. Conversely, the user can look up a programme online and then 'push' relevant information to the connected television in order to tune it to the right channel" (Vinayagamoorthy and Kramskoy, cited in Züll and Mikelić Preradović, 2013, p. 129).

Broadcasters may often and regularly post unique emotional content with an exactly determined time schedule and pre-planned posts for weekends and non-business hours. It is

necessary to have content strategy and to plan (real-time) tweets along the broadcast of the television program by celebrities (as they have huge online audiences), competition judges, or have other experts include countdown posts, for example about exclusive behind the scenes material (backstage, interviews) which is not included in the broadcast. It is necessary to keep the audience and maintain the engagement, to ask questions (news and entertainment and current affairs shows), to let them vote (game and talent shows) and to generally encourage the audience to support the show. However, the leisure time of the social media team is the time when audiences are most active. It is necessary to schedule the volume of social media activity including the number of posts for each platform. As social television relates to the communicative exchange about linear television content (Buschow, Schneider and Ueberheide, 2014) it needs be structured that pre-communication in pre-broadcasting phase and follow-up communication in post-broadcasting phase are included.

Figure 6.2. Timeline of social media postings

The goal is to keep the audience in the created horizontal network, to direct the traffic within this network and to get audiences back to the television screen. Social media sentiment (4.4 and 5.1) towards a television program can be shown during the broadcast. The social response of television audiences could allow the audience to influence or to participate in the television program.

Despite the vast potential that social media brings, the television industry still seems focused on social media activity primarily as a one-way promotional channel, and they need to find ways of analysing consumers' conversations and turn the information into valuable insights. Some television hosts have incorporated Twitter content into the core of their broadcasts, actively including people's comments on air and responding directly to tweets

during their programs (Neufeld, 2011). “A recent study by Twitter closely monitored the real-time ripple effects on Twitter when the television host directly asked their audience to participate with them via Twitter during the premier of a new series. The amount of mentions and follows hovered at around one hundred per minute for the duration of the broadcast with the occasional spike at moments of interest, with an overall massive increase in traffic compared to normal non-broadcast hours” (Neufeld cited in Züll, Mikelić Preradović, Boras, 2013, p. 281).

It seems that some programs are more suitable than others for communicating and sharing (Chorianopoulos and Lekako, 2008). It can be assumed that for some programs, considered to be social television programs, more communication can be found on the internet than for others.

The broadcaster (Züll, Boras and Mikelić Preradović, 2013) sets up, registers and engage social media accounts. The accounts are maintained in the name of the public institution, though not for personal use. Every activity is performed in the name of the programs, teams or brands and has a quality level of communication in respect to language and content. The social media activities labelled ‘official’ have the same status as mainstream broadcast. The number of accounts (Twitter or Facebook), are controlled by the senior editors and producers of the relevant department aiming for a coherent, stable and distinctive voices of television broadcasters. The competence and the satisfaction of the employees and the image of the television channel to the outside media world would be guaranteed.

A social media guide (BBC, 2011) ensures that the ‘official accounts’ are official and not personal and the conversation is led by an official voice. In general, the social media accounts are an instrument for the broadcaster to distribute content, therefore the question regarding ‘who is running the official account’ can be raised. This is of high interest particularly when the person behind the account changes positions or television stations. The official accounts listed on the channel website should be in line with the channel guidelines for the use of social networking. Social media is good writing, photography, and video in a shared space with the audience. Photos and videos need to be checked for authenticity and copyright purposes (Rieber, 1994). As social media is originally designed for people and not public institutions it is important that public institutions define the line between personal and professional as a misjudgement in communication could undermine the credibility of the television company.

All activities need the branding of the television channel and activities need to be consistent with the channel social media strategy. To the extent possible all social media contributions need be checked by an editorial department. The need for care applies all the time, because anything said via social media is official.

The need for new job profiles to be created arises out of the need to establish multi-social communication systems within the broadcaster. All of these new tasks ask for interdisciplinary knowledge: technological background, knowledge of mobile business, IT-understanding, writing skills, creative background, copyrighting and licensing knowledge (Habajec, 2014).

Social Media Manager	Videographers/Photographers
Social Media Producer	Livestream Manager
Application Developer	Social Media Buyer
Content Editors/Creative Writers	Digital Public Relation Manager

Figure 6.1. Social media creates jobs in television

Usually, broadcasters have professional employees, and the technical and administrative support staff is part of the mix of ongoing innovation teams and projects rather than being placed in separate departments as they develop in a complex, rapidly changing environment (Daft, 2012).

A *social media manager* (Bottles and Sherlock, 2011) as head of a team needs to coordinate all social media activities and to install a controlling system to supervise uploads and posts by the broadcaster and the audience. A *social media producer* is in charge of the production of social media content for social media channels parallel to the production of content for television. They are also in charge of copyright and rights clearance for all content posted on social media channels. An *application developer* creates, for example, voting applications for second screen usage to keep the audience within the convergent world (chapter 2) of the television program and to get the live audience sentiment which enables online audience research. Content editors/creative writers are experienced internet writers with online experience; they create interaction with the audience, encourage the audience to engage and to show their support, and they also inform the audience. *Videographers/Photographers* supply

high quality (live stream) videos and clips for social media channels. A *livestream manager* plans and pipes live video feeds onto the internet. *Social media buyers* purchase strategic marketing and media reach (on platforms such as Facebook, Twitter, and YouTube).

Measuring the social media effort and evaluating the value and return of investment in the short and longer-term remains a challenge (Lovreček, 2014). Recent research has investigated the impact of social media strategies of television broadcasters on viewer engagement (Proulx and Shepatin, 2012; Hill, Benton and Peng, 2012). Proulx and Shepatin recommend, for example, using social television ratings analytics tools to find and target active audiences. Nielsen announced its television ratings in partnership with Twitter in 2012 positioning it as a tool that helps to understand how television audiences respond to a television program. In 2014 Nielsen announced that they expanded the information for its Twitter television ratings service with the launch of demographic data to allow the industry and researchers to identify the age and gender of those who are tweeting about various television shows and events in the United States (Perez, 2014). Overall, the internet offers a massive number of tools and a huge amount of audience behaviour data to be researched, while keeping in mind that scholars should be aware of ethical, legal and technical limitations (Züll, 2014).

7 CONCLUSION

The idea of this work was to investigate the phenomenon of increasing social media usage and activities in the context of free-to-air television. New technologies and in particular the internet brought change and transformation to the media environment, therefore chapter 2 started with describing the change of television media. The term *internet* expresses a spectrum of manifold meanings, technical currents and developments which enable communication and interactivity of users as it forms the infrastructure as a precondition for networking. Internet describes technology and opportunities of usage at the same time. It is the basis of change also because it is a way of transmission of television signals and therefore a driving factor for fragmentation of television markets. Therefore, a paragraph at the beginning of this work is dedicated to the internet. The internet can be considered a trigger of multimedial change of television and television viewers.

Starting with the television landscape, the internet and digitisation of transmission channels leads to increasing fragmented television markets with a growing number of television (niche) channels. This could be observed in Croatia as in all other television markets in Europe and across borders. The internet is a form of transmission of television signals gained in connection with wireless interconnectivity and development of mobile devices such as mobile phones, tablets and laptops. Mobile devices are used to watch television content; if they are used parallel to television consumption (at home) they are often called second screens. The usage of mobile devices in combination with the internet has changed habits and behaviour of television viewers as it enables the user to watch (linear or non-linear) television content via a diverse offer of television and video applications. Furthermore, mobile devices facilitate and accelerate the usage of social media applications such as Facebook or Twitter parallel to television consumption, offering the technical infrastructure to exchange any kind of information such as texts, photos, links and videos.

In their definition of social media, Haenlein and Kaplan (2010) focus on user-generated content but in fact mega-publishers (Levine, 2012) such as traditional media dominate the content of social media which people discuss, share or forward (chapter 4). The business model of social media providers is the access to user data. It seems that computational knowledge is a main requirement for the human scientist to be able to use social media data for analysis. The question was asked if the information could be a benefit compared to traditional television audience research approaches in particular to television ratings research based on

people-meter-system methodology, as this dominates the daily business of television makers. For this reason, chapter 3 investigated the traditional audience research and evaluated the significance of this currency to the television industry. Although discussed and often criticised by industry and sciences, the methodology based on people-meter-systems proved to be an appropriate instrument to process and generate data and to submit a detailed and complete picture on a minute by minute basis 365 days a year which, for a long-term perspective, one can expect will not be substituted.

Furthermore, the analysis of this work showed that one cannot simply reduce television ratings research to a numbers-based approach. The people-meter-system is extremely informative, particularly because it has statistical relevance and combines quantitative information of panel members with ratings numbers (3.2.2). Two arguments are motivation to still continue (or more than ever) considering and dealing with social media analytics as an instrument to learn more about television audiences: first the question of whether or not increasing fragmented television markets and mobile non-linear television consumption reduce the importance of traditional television ratings research in the long term, and therefore new and more relevant methods are needed in the future. Second, television ratings research is marked with the significant blemish of not submitting information about the opinion of television viewers towards a show, and that it is simplified by equating high ratings with positive opinions which, leading to the attraction of social media analytics where the researcher can find data about opinions towards any kind of products or media content.

The detection of opinions of television viewers towards a television program is therefore attractive for program makers and producers. Social media data as a tool for market research is therefore a trend topic within the industry. From the beginning, it was obvious that the analysis of social media data cannot submit a complete or representative picture of television audiences as only a part of the population, primarily younger people, use social media applications to exchange and share opinion about media content. This problem is increased by the fact that not every television program is equally suitable for social media. Only strong television brands motivate the television viewers to respond on social media and to communicate about their viewing experiences.

This work tried to find a systematic way and satisfactory solution to the question of how to approach social media data (chapter 4.3) and to answer the question of how the results

could eventually complement information which can be gained through traditional audience research (chapter 6).

The choice of Twitter as a data basis and source was driven by the fact that Twitter is known for its (more) open data policy compared to other platforms as it allows the user in general to access data and even more to retrieve data through the Twitter application programming interface (API). Twitter data as a source within the framework of this research should only be understood as an example for a data source. In general, developed methods to analyse social media data should be applicable to other social media data sources as well. If Facebook would make its data available to the researcher it could be attractive as the usage numbers are higher compared to Twitter, particular in Croatia. As Facebook is not a microblogging service but offers manifold tools it could improve the analysis. After the selection of the data source (chapter 4.1) a methodology had to be found to approach and detect data related to television (chapter 4.3).

Opinion mining tools as an instrument to approach data was selected and evaluated as an approach to find data related to television consumption (Züll and Mikelić Preradović, 2013) as many cost-free tools can be found on the internet. The number of social media analysis tools is growing and changing but valid tools which the researcher has free access to are rare. Transparency about their functionality and algorithms are in general not given.

Different television shows (serial and non-serial) were randomly chosen and the title of the shows was used as a search term and applied to a selection of opinion mining tools. The tools taught the researcher which queries deliver results and which will not and the critical point was the definition of the search term. Only if tools show relevant hashtags, further research can bring additional results. Data results were compared and evaluated and the functionalities of the tools were investigated. The question regarding which tool submits accurate results and the evaluation of the algorithm used by the individual tools was not given a satisfactory response. The most significant issue for this part of the research was that all tools delivered different results, although all queries were made in a selected time frame.

It was considered as a weakness of the opinion mining tools that the search queries only submit snapshots of the data situation. For that reason, Yourtwapperkeeper was introduced as an additional tool to retrieve and export data for a longer time frame based on key word and hashtag research. This tool proved to be convenient and in particular the possibilities to work with and export data proved to be helpful. In particular, the metadata set which comes along

with the tweet text submits additional information such as when (time), where (geolocation) and from whom (IP address) the tweet was generated. Again, the combination of different data provides valuable information.

In the context of metadata sets privacy issues of users have to be taken into consideration (Züll, 2013; Züll, 2014) and the anonymity of traditional television ratings audience research was missed. In regards to algorithms and sources of data Yourtwapperkeeper is not open. Tools and applications only have an exemplary character as they often transform into commercial offers and after a certain time frame the software is no longer available at to the researcher no costs.

If hashtags are introduced by the broadcaster as part of a promotion strategy to interact with the audience, the discussion on Twitter can be located more easily (Kojundžić, 2014). Usually a group of hashtag/keywords – a so-called word cloud - exists around a television program such as the title of the program or the main cast or titles of episodes. This means that defining appropriate queries represents a challenge. How to find the best Twitter acronyms related to a television program and is there a need and possibility to tie video streams to the social media comments? How can the user find the discussion related to a certain program?

As a specific circumstance in Croatia, it was observed that television broadcasters in Croatia rarely perform support communication on Twitter by broadcasting hashtags or keywords to inspire the audience to talk about their shows on Twitter.

As a result of the fact that algorithms of data retrieve and data sources of publicly available opinion mining tools and software such as *Yourtwapperkeeper* seemed not to show sufficient transparency from a research perspective, the approach of considering Twitter for data collection or as a library was analysed (chapter 4.3.2).

This idea was inspired by the press release of the US Library of Congress in 2013 in cooperation with Twitter, when they published the agreement that Twitter would give all its data to the public library to make it available for public research. This press release was considered as relevant to this work and is therefore attached to this work as appendix 3, as it interprets Twitter data as a public library of data which should be accessible for public research. One could argue that user-generated social media data should be owned by users, for example, as a public data collection (or a public good). The opposite happens when social media providers develop profitable business models based on user data as currency (Züll, 2013).

It is the idea of chapter 4.3.2 to develop an approach for having full access to data and no limitation or filter by any kind of interposed tool, application or company. Furthermore, the idea to work with a data corpus was considered as attractive, given that the question of how to locate communication if people do not use special keywords introduced by broadcasters to organise their communication was not answered. It is a fact that no comparable agreement between Twitter and any public library outside the United States exists. Lomberg (2012) described data corpuses as useful method to study communication practices of users as it allows the research to retrieve and harvest internet data, which means Twitter data and relevant Meta data such as time stamp, geolocation or relevant user information.

In search of a response to the issue of how to access Twitter data, the Twitter data grant program released by Twitter was found in 2014. Within the framework of this research the author of this thesis applied (chapter 5.2 and appendix 1 and appendix 2) to receive the data grant by Twitter with the purpose of analysing the public sphere in Croatia, Serbia and Bosnia and Herzegovina and to search for conversation on Twitter in the context of relevant political, media and other topics. Twitter received more than 1,000 applications from academic research groups worldwide and the Croatian application was not taken into consideration.

Furthermore, in 2014, Ljubešić, Fišer and Erjavec published the results of their research about data Twitter collections for smaller languages (chapter 4.3.2). The research is based on the application of an open-source tool called TweetCat which was designed to collect tweets and build corpora for smaller languages. Two web corpora were build, one for Croatian and Serbian, and one for Slovene, using seed terms and simple language identification modules. The tool did not discriminate between Croatian and Serbian tweets, thanks to the complexity of the task. The Croatian and Serbian web corpora (Ljubešić, Fišer and Erjavec, 2014) was collected over a time frame of 235 days.

Any Twitter data collection (corpus) can only be a (demographic) extract of the communication of television audiences. As an idea for finding data with informative value in a systematic manner, the method of sampling (chapter 5.2) was applied.

It was possible to extract a ranking list of television related tweets and to find keywords relevant to television (chapter 5.2). It was discovered that most tweets were related to Serbian media, some international television shows could be discovered and Croatian and Serbian broadcasters could be identified. It was discovered that most of the tweets were created by the television channels themselves with the purpose of promoting the series. The tweets mostly

contained information about a serial program or announcement of the program. Only a minority of tweets found could be used to provide insights into the communication activities of audiences.

It was confirmed that different genres have different levels of suitability for social media activity and therefore different levels of suitability for research and social media analysis (Züll and Mikelić Preradović, 2013). Live, linear television programing seems to be more suitable for analysis, and the more "social" a television program is, the more comments on Twitter that can be found. International television shows are suitable for analysis, but the researcher needs to be aware that the research needs to be designed around and adapted to the television broadcast in the individual country.

Serial one-off events (such as the Oscars) need to draw huge public attention (on Twitter) and need to be watched by a large global audience to be suitable for analysis. More examples could be the annual Eurovision song contest which attracts hundreds of millions of viewers, or royal events such as weddings.

As a result, it can be stated that the way of working with data corpus and sampling could be considered a possible approach to finding relevant keywords. The vital precondition is that broadcasters need to inspire and create communication around their programs in the social media world to maintain a relationship with their younger audiences, for example as has been shown by the BBC in the United Kingdom (Züll, Boras and Mikelić Preradović, 2013).

As a result one can say that television broadcasters are well-advised to develop their social media strategy (chapter 6.3) to develop and shape their relationship particularly with the younger audience (Bulkeley, 2010). In Croatia and many other European countries this has not yet been the case in a sufficient amount of activities. Therefore, and as a result of this research work, a recommendation for broadcasters in Croatia is to develop and integrate social media activities in their daily routines to create interactivity and communication around their television programs.

The precondition for analysis of social media data is that digital traces (Züll, 2013) can be connected back to the television broadcaster. When considering news, sporting events and elections it can be difficult to assign the comments found on the internet to the broadcast of the program, as the comments could be assigned to the event itself. It is especially important to make this distinction when researching the behaviour of television audiences and to know the

reason or impulse behind the user posting the comment. This question also arises when people watch non-linear television and post time-shifted comments related to the original broadcast. However, it is a fact that social media activity in the context of free-to-air television is the exception for strong television brands and topics rather than the rule, and that critical mass is needed to show relevance.

Ideally, there would be a connection between the program the user is watching and the companion device such as mobile phone or tablet. For social media analysis it is important that the referring online discussion can be assigned to the program and the broadcaster. Ideally, a virtual lounge room exists.

8 BIBLIOGRAPHY

Acar, A. and Muraki, Y. (2011). "Twitter and natural disasters: Crisis communication lessons from the Japan tsunami", 15th April, *International Journal of Web Based Communities*, 7(3), pp. 392-402. Inderscience Publishers. Available at: www.sciencedaily.com/releases/2011/04/110415154734.htm (Accessed: 15th April 2011).

Agency for Electronic Media of the Republic of Croatia (2016) Agencija za elektronicke medije. Available at: http://www.e-mediji.hr/hr/ (Accessed: 7th August 2016).

Agrawal, R., Rajagopalan, S., Srikant R., and Xu, Y. (2003) "Mining newsgroups using networks arising from social behaviour", in *Proceedings of WWW*, pp. 529–535.

Alexa (2016) *Top Sites in Croatia*. Available at: http://www.alexa.com (Accessed: 19th March 2016).

Allin1Social (2016) *Facebook Statistics for Croatia*. Available at: http://www.allin1social.com/facebook-statistics/countries/croatia (Accessed: 9th April 2016).

Andrejevic, M. (2008) "Watching television without pity: The productivity of online fans", *Television & New Media (TVNM)*, 1(9), pp. 24-46.

Ang, I. (1991) *Desperately Seeking the Audience*, London: Routledge.

Arndt, R.Z. (2011) *Bluefin mines social media to improve TV analytics*, *Fast Company* (2011, November 7). Available at: http://www.fastcompany.com/node/1793473/print (Accessed: 20th May 2016).

Artemi, C. (2005) "A multidimensional model of an electorate", in *Proceedings of the 4th WSEAS International Conference on Non-linear Analysis, Non-linear Systems and Chaos* (NOLASC '05), Mladenov, V., Bardis, N. and Damgov, V.(eds), World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, pp.68-71.

Baccianella, S., Esuli A. and Sebastiani F. (2010) "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining", *Proceedings of LREC*, pp. 2200–2204.

Bakos, G., 2013. *KNIME essentials*. Packt Publishing Ltd.

Barnard, S.R. (2012) *Twitter and the Journalistic Field: How the growth of a new(s) medium is transforming journalism*, PhD thesis. University of Missouri.

Bautin, M., Vijayarenu, L. and Skiena, S. (2008) "International sentiment analysis for news and blogs", in *Proceedings of the International Conference on Weblogs and Social Media* (ICWSM).

BBC News Group (2011) *Social Media Guidance for staff*. Available at: news.Bbc.co.uk., pp. 1-3
(Accessed: 16th May 2016).

Beheshti-Kashi, S. and Makki, B. (2013) “The use of social software in the context of daily news”, in *Recent Advances in Knowledge Engineering and Systems Science* WSEAS Press, pp. 225-229.

Benamara F., Cesarano C., Picariello A., Reforgiato D., and Subrahmanian, V.S. (2007) “Sentiment analysis: Adjectives and adverbs are better than adjectives alone”, in *Proceedings of the International Conference on Weblogs and Social Media* (ICWSM).

Benhardus, J. and Kalita, J. (2013) “Streaming Trend Detection in Twitter”. *International Journal of Web Based Communities*, *9*(1), pp.122-139.

Bennett, S. (2014). “What (And Why) Do People Share on Social Networks?”, Social Times (15th July 2014) [Online]. Available at: www.adweek.com (Accessed: 26th April 2016).

Bens de, E., Golding, P. and McQuail, D. (eds) (2005) *Communication Theory & Research – An EJC Anthology.* London, Sage Publications.

Bergstrom, T., Blume, L., Varian, H. (1986) “On the private provision of public goods”, *Journal of Public Economics*, 29(1).

Benhardus, J. and Kalita, J. (2013) “Streaming trend detection in Twitter”, *International Journal of Web Based Communities*, 9(1), pp.122-139. Available at: https://www.semanticscholar.org/paper/Streaming-trend-detection-in-Twitter-Benhardus-Kalita/9d73edb3891e0ce99af738b68b7a857a52db4a84/pdf (Accessed: 15th March 2015).

Berger, A.A. (2015) Media and communication research methods: An introduction to *qualitative and quantitative approaches*. Sage Publications.

Berry, D. M. (2004) "Internet Research: Privacy, Ethics and Alienation: An Open Source Approach", *Internet Research.*14 (4), pp. 323-332.

Best, S. and Breunig, C. (2011) „Parallele und exclusive Mediennutzung", *Media Perspektiven,* 1/2011, pp. 16-35.

Billsus, D., Pazzani, M.J. (1999) "A personal news agent that talks, learns and explains", Proceedings of the Third International Conference on Autonomous Agents, pages 268–275. ACM Press, pp. 268-275. Available at: http://www.cs.rutgers.edu/~pazzani/Publications/agents99-news.pdf (Accessed: 15th August 2015).

Bjur, J. (2009) *Transforming audiences - Patterns of Individualization in Television Viewing*. PhD thesis. University of Gothenburg.

Bottles, K. and Sherlock, T., 2011. Who should manage your social media strategy? *Physician Executive*, *37*(2), pp.68-72.

Bourdon, J. and Méadel, C. (2011) "Inside Television Audience Measurement: Destructing the ratings machine", *Media, Culture & Society,* Sage, pp.1-10.

Boyd, D. and Crawford, K. (2012) "Critical Questions for Big Data", *Communication & Society*, 15(5), pp.662-679.

Boyd, D., Golder, S. and Lotan, G. (2010) “Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter”, *HICSS-43. IEEE: Kauai, HI*. January 6.

Bredl, K., Hünniger J. and Jensen J.L. (2012) “Methods of Analyzing Social Media – Introduction to the Special Issue”, *Journal of Technology in Human Services,* 30(3-4), pp. 141-144.

Bredl, K., Ketzer, C., Hünniger, J. and Fleischer, J. (2014) “Twitter and Social TV: Microblogging as a new approach to Audience Research”, in Patriarche, G., Bilandzic, H., Jensen, J.L. and, Jurisic, J. (eds) *Audience Research Methodologies – Between Innovation and Consolidation*. Routledge: New York and London, pp.196-211.

Brügger, N. (2011) *Web archiving-Between past, present, and future*, in Consalvo, M. and Ess, C.M. The Blackwell Handbook of Internet Studies, pp. 190-212. Oxford, UK: Wiley-Blackwell.

Brügger, N. (2005) *Web archiving*. Aarhus, Denmark: Centre of Internet Research, Aarhus University.

Bruns, A., and Liang, E. (2012) “Tools and methods for capturing Twitter data during natural disasters”, 2 April*, 'First Monday [Online],* 17(4). Available at: http://firstmonday.org/article/view/3937/3193
(Accessed: 5th May 2014).

Bruns, A., and Stieglitz, S. (2013) “Towards More Systematic Twitter Analysis: Metrics for Tweeting Activities”, *International Journal of Social Research Methodology,* 16(2): 91-108.

Bruns, A. (2012) *Twapperkeeper and Beyond: A Reminder, 9 January.* Available at: http://mappingonlinepublics.net/2012/01/09/twapperkeeper-and-beyond-a-reminder/ (Accessed: 9th June 2012).

Bruns, A., Liang, Y. E. (2012) "Tools and methods for capturing Twitter data during natural disasters", 2nd April, *First Monday* [online]. 7(4). Available at: http://pear.accc.uic.edu/ojs/index.php/fm/article/view/3937/3193 (Accessed: 29th April 2012).

Bruns, A. (2011) *Switching from Twapperkeeper to Yourtwapperkeeper*, 21 June. Available at: http://mappingonlinepublics.net/2011/06/21/switching-from-twapperkeeper-to-yourtwapperkeeper/. (Accessed: 25th June 2013).

Bruns, A., and Stieglitz, S. (2012) "Quantitative Approaches to Comparing Communication Patterns on Twitter", *Journal of Technology in Human Services*, 30(3-4), pp. 160-185.

Bertelsmann Stiftung (2008) *Croatia Country Report, Bertelsmann Transformation Index Country Report,* Gütersloh.

Burgess, J. and Bruns, A. (2012) "Twitter Archives and the Challenges of ´Big Social Data for Media and Communication Research", *M/C Journal a Journal for Media and Culture* [online], 15(5).

Bulkeley, W. (2010) "Relying on Relationships to Rebuild TV Audiences", Social TV, Technology Review, (5/6), pp. 55-56.

Buschow, C., Schneider, B. and Ueberheide, S., (2014) “Tweeting television: Exploring communication activities on Twitter while watching TV”, *Communications-The European Journal of Communication Research*, 39(2), pp.129-149.

Buzzard, K. (2012) *Audience Ratings in a Digital Age – The Ratings Industry from analog to digital*. [e-book-reader]. Routledge: New York.

Buzzard, K. (2002) “The peoplemeter wars: A case study of technological innovation and diffusion in the ratings industry”, *Journal of Media Economics*, 15(4), pp. 273 – 291.

Cambria E., Havasi C. and Hussain A. (2012) “SenticNet 2: A semantic and affective resource for opinion mining and sentiment analysis”, *Proceedings of AAAI FLAIRS,* pp.202–207.

Car V. (2010) “Television in New Media Environment”, *Media Studies* 1(1–2), pp.91–104.

Carbonell, J.G. (1979) *Subjective Understanding: Computer Models of Belief Systems*, PhD thesis, University Yale.

Cardie, C., Wiebe, J., Wilson, T. and Litman, D. (2003) “Combining low-level and summary representations of opinions for multi-perspective question answering”, *Proceedings of the AAAI Spring Symposium on New Directions in Question Answering*, pp.20–27.

Cardoso, G. (2008) “From Mass to Network Communication: Communicational models and the Informational Society”, *International Journal of Communication*, 2, pp. 587-630.

Cardoso, G. (2006) *The Media in the Network Society: Browsing, News, Filters and Citizenship*, librosmetabiblioteca [Online]. Lisbon, Portugal: CIES – Centre for Research and Studies in Sociology. Available at: http://libros.metabiblioteca.org/handle/001/195 (Accessed: 7th August 2016).

Carpentier, N., Schrøder, K.C., Hallett, L. (2014) “Audience/ Society Transformations”, in Carpentier, N., Schrøder, K.C. and Hallett, L. (Eds) *Audience Transformations Shifting Audience Positions in Late Modernity.* Routledge: New York, pp.1-12.

Car, V. and Andrijašević, I. (2012), “Mapping Digital Media: Croatia”, Dragomir, M., Thompson, M., and Walter, M. (eds), Open Society Foundations Publication. Available at: https://www.opensocietyfoundations.org/reports/mapping-digital-media-croatia (Accessed: March 2016).

Castells, M. (2015) *Networks of outrage and hope – Social Movements in the Internet Age*. Polity Press: UK.

Castells, M. (2009) *The Power of Identity: The Information Age. Economy Society and Culture*. Volume II. 2nd edn. Oxford: Wiley Blackwell.

Castells, M. (2007) “Communication, Power and Counter-power in the Network Society”, *International Journal of Communication*, 1, pp. 238-266.

Castells, M. (2000) *The Information Age – Economy, Society, and Culture – Volume I The rise of the Network Soc*iety. 2nd edn. Oxford: Wiley Blackwell.

Chiribuca, D, I. Pah and D. Hunyadi (2008) "Social and cultural challenges of the new communication technology used in education oriented activities", *Recent Advances in Education and Educational Technology*, WSEAS Press, pp. 129-134.

Cochran, W.G. (1977) Sampling Techniques, 3rd edn. New York: Wiley.

Chorianopoulos, K., and Lekakos, G. (2008) "Introduction to social TV: Enhancing the shared experience with interactive TV", *International Journal of Human-Computer Interaction*, *24* (2), pp. 113-120.

COST Action ISO 906 (2010-2014) "Transforming Audiences, transforming societies", Available at: http://www.cost-transforming-audiences.eu.

Courtois, C. and Mechant, P. (2014) "An Evaluation of the Potential of Web 2.0 APIs for Social Research", in Patriarche, G., Bilandzic, H., Jensen, J.L., Jurišić J., (eds) *Audience Research Methodologies – Between Innovation and Consolidation*. Routledge: New York and London, pp. 212-224.

Croatia week (2014) "Who wants to be a millionaire court case drags on for Croat" (8 September 2014). Available at: http://www.croatiaweek.com/who-wants-to-be-a-millionaire-court-case-drags-on-for-croat/
(Accessed: 7th October 2014).

Croatian State Bureau of Statistics (2016). Državni zavod za statistiku. Available at: http://www.dzs.hr
(Accessed: 7th August 2016).

Čular, G. (2000) *Political Development in Croatia 1990-2000: Fast Transition – Postponed Consolidation Political Development in Croatia* (Politička misao). Vol. XXXVII, No. 5: 30–46.

Danesi, M. (2002) *Understanding Media Semiotics*. Oxford University Press.

Das, M., Ester, P. and Kaczmirek (eds) (2011) *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*. [Online] Routledge: New York: Taylor & Francis. Available at: http://www.phd.meghan-smith.com/wp-content/uploads/2015/07/Smyth-and-Pearson.pdf
(Accessed: 7th August 2016).

Das, S., Chen, M. (2001) "Yahoo! for Amazon: Extracting market sentiment from stock message boards", *Proceedings of the Asia Pacific Finance Association Annual Conference* (APFA).

Dave, K., Lawrence, S., Pennock, D.M. (2003) "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews". *WWW 2003*, May 20-24, 2003, Budapest. Available at: http://kushaldave.com/p451-dave.pdf
(Accessed: 6th August 2016).

Deller, R. (2011) "Twittering on: Audience research and participation using Twitter", in *Participations - Journal of Audience and Reception Studies*, 8(1). Available at: http://www.participations.org/Volume%208/Issue%201/PDF/deller.pdf (Accessed: 15th January 2013).

Dijck van, J. (2013) *The Culture of Connectivity: A critical History of Social Media*, Oxford: University Press.

Dijk van; J. (2013) *The Network Society*. 3rd edition. University of Twente. Sage Publications.

Dini; L. and Mazzini G. (2002) "Opinion classification through information extraction", *Proceedings of the Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields (Data Mining)*, pp. 299–310.

Doughty, M., Rowland D., & Lawson, S. (2011) Co-viewing with digital backchannel streams, *Proceedings EuroITV '11 Proceedings of the 9th international interactive conference on Interactive television*, pp.141-144.

Daft, R., 2012. Organization theory and design. Nelson Education.

Dugan, L. (2012) "The Oscars on Twitter: 18,718 Tweets-Per-Second Record, But Dwarfed By Grammys" (27th February 2012), *Social Times* [Online]. Available at: http://www.adweek.com/socialtimes/the-oscars-on-twitter-18718-tweets-per-second-record-but-dwarfed-by-grammys/460489 (Accessed: 15th March 2012).

Ducheneaut, N., Moore, R.J., Oehlberg, L., Thornton, J.D. and Nickell, E. (2008), "Social TV: Designing for Distributed, Social Television Viewing", *International Journal of Human-Computer Interaction*, 24(2), pp. 136-154. Available at: http://www.parc.com/publication/2123/socialtv.html
(Accessed: 12th January 2013).

Duggan, M., Ellision, N.B., Lampe, C., Lenhart and Madden, M. (2015) "Social Media update 2014", 9 January 2015, Available at: http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf
(Accessed: 15th January 2015).

Dzitac, I. and Moisil, I. (2008) "Advanced AI Techniques for Web Mining", *Proceedings of 10th WSEAS International Conference on Mathematical Methods, Computational Techniques and Intelligent Systems (MAMECTIS '08)*, Corfu, Greece, pp. 343-346.

Edelsburg, N. (2011) "Social Guide set to launch 'Intelligence' analytics tool". *Lost Remote* [Online], 1 November 2011. Available at: http://www.adweek.com/lostremote/social-guide-set-to-launch-intelligence-analytics-tool/23489
(Accessed: 31st March 2012).

European Commission Union Law (2001) "Communication from the Commission on the application of State aid rules to public service broadcasting", *Official Journal of the European Communities,* 15 November 2001, C 320, 44. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:C:2001:320:TOC
(Accessed: 8th August 2016).

Fiske, J. and Hartley, J. (2003) Reading Television, 2nd edition. Routledge: London (Mehuen and Co. Ltd., 1st edition).

Freedom House (2015) "Freedom of the Press Croatia". Available at: https://freedomhouse.org/report/freedom-press/2015/croatia
(Accessed: 15th March 2013).

Friedman, W. (2012) "Social media buzz ups traditional TV ratings", *Media Daily News 14 March 2012* [Online]. Available: http://www.mediapost.com/publications/article/170201/social-media-buzz-ups-trad-tv-ratings.html
(Accessed: 24th September 2013).

Friedman, W. (2009) "Q Scores tout 'CSI', 'Event'. ", *Media Daily News*, 9 February 2009 [Online]. Available: http://www.mediapost.com/publications/article/144633/
(Accessed: 9th February 2009).

Fukuhara, T., Nakagawa, H. and Nishida, T. (2007), "Understanding sentiment of people from news articles: Temporal sentiment analysis of social events", in *Proceedings of the International Conference on Weblogs and Social Media* (ICWSM).

Futurescape (2012) "Strategy Reports on the Future of Television 2012". Available at: http://www.futurescape.tv/
(Accessed: 15th March 2013),

Garside, J. (2015), "Twitter puts trillions of tweets up for sale to data miners", *The Guardian*, 18th March 2015 [Online]. Available at: theguardian.com.
(Accessed: 23rd April 2016).

Gitlin, T. (2000 [1983]) “Inside Prime Time”. University of California Press.

Giglietto, F. and Rossi, L. (2013) “The open laboratory: Limits and Possibilities of Using Facebook, Twitter, and YouTube as a Research Data Source”, in Journal of Technology in Human Services.” 30, 3-4: 145-159.

Go, A., Bhayani, R. and Huang, L., (2009) “Twitter sentiment classification using distant supervision”, *CS224N Project Report*, Stanford, 1, p.12.

Department for Culture, Media and Sport (2006) “Broadcasting Copy of Royal Charter for the continuance of the British Broadcasting Corporation”. Available at: http://www.bbc.co.uk
(Accessed: 26th May 2012).

Garton, L., Haythornthwaite, C., Wellmann, B. (1997) “Studying On-Line Social Networks”, *Journal of Computer Mediated Communication* (JCMC), 3(1).

Go, A., Bhayani, R. and Huang, L. (2009) “Twitter Sentiment Classification using Distant Supervision”, Technical report, Stanford Digital Library Technologies Project.

Godbole, N., Srinivasaiah, M., and Skiena, S. (2007), “Large-scale sentiment analysis for news and blogs”, in *Proceedings of the International Conference on Weblogs and Social Media* (ICWSM).

Goldstein, I. (2010) “Dvadeset Godina Samostalne Hrvatske” (Twenty Years of Independent Croatia), Zagreb: Novi Liber.

Government of the Republic of Croatia (2015) “Government of Croatia in 2015 named as the most prolific government on Twitter in Europe”. Available at: https://vlada.gov.hr/government-of-croatia-in-2015-named-as-the-most-prolific-government-on-twitter-in-europe/16873
(Accessed: 16th April 2016).

Green, L. (2001) “Communities of broadcasting and communities of interactivity”, in *Communication World*, WSEAS Press: pp. 158-163.

Hasebrink, U. and Popp, J. (2006) “Media Repertoires as a result of Selective Media Use: A Conceptual Approach to the Analysis of Patterns of Exposure”, in *Communications*, 31 (2): pp. 369-387.

Harrington, S., Highfield, T., Bruns, A. (2013a) “Twitter as a Technology for Audiencing and Fandom: The #Eurovision Phenomenon”, in *Information, Communication & Society´.16* (3): 315-339.

Harrington, S., Highfield and T., Bruns, A. (2013b) “More than a backchannel: Twitter and Television”, in *Journal of Audience Reception Studies,* Volume 10, Issue 1: 405-409.

Hesmondhalgh, D. (2007) The Cultural Industries. London: Sage.

Highfield, T. (2012) “Talking of Many Things: Using Topical Networks to Study Discussions in Social Media”, in *Journal of Technology and Human Services*, vol. 30, no. 3-4: 204-218.

Highfield, T., Kirchhoff, L., and Nicolai, T. (2011) “Challenges of tracking topical discussions online”, *Social Science Computer Review*, 29 (3), pp. 340-353.

Hill, S., Benton A.D., and Peng, J. (2012) “Social TV: Analyzing the Impact to of Social Media Strategies on Viewer Engagement”, 33rd International Conference on Information Systems, Orlando, Florida, 16-19 December. Available at: http://wan.poly.edu/KDD2012/forms/workshop/SOMA2012/doc/wks_submission_3.pdf (Accessed: 4 May 2014).

Hitchcock S. (2013) “Collecting and archiving tweets: a Data Pool case study”, University of Southampton. Available at: http://eprints.soton.ac.uk (Accessed: 30th June 2013).

Holmes, D., (1997) “Virtual identity: communities of broadcast, communities of interactivity”, in *Virtual politics: identity and community in cyberspace.* Sage. London: 26-45.

[hootsuite.com] Hootsuite.

Howell, R. (2005) “Visual Culture”. Oxford, Polity.

Hrastovčak, T. (2013) “Hrvati preferiraju Facebook, Ali popularnost Twittera raste”, in *Poslovni Dnevni* (15th February 2014). Available at: http://www.poslovni.hr/ (Accessed: 21st January 2014).

Hrvatska Radiotelevizija (2009) “The Evolution of Croatian Radio and Television”, 1 January 2009. Available at: http://www.hrt.hr/23183/povijest/the-evolution-of-croatian-radio-and-television
(Accessed: 1st April 2014).

Internet World Stats (2015) “Usage and Population Statistics”. Available at: http://www.internetworldstats.com/stats.htm.
(Accessed: 1st April 2015).

Haenlein M. and Kaplan, A. M. (2010) “Users of the world, unite! The challenges and opportunities of Social Media”, in *Business Horizons*, 53(1): 59-68.

Jansen B. J., Zhang M., Sobel K., Chowdury A. (2009) “Twitter power: Tweets as electronic word of mouth”, in *Journal of the American Society for Information Science and Technology*, vol. 60, No. 11: 2169–2188.

Jennes, Iris and Pierson, Jo (2012), “Audience measurement and digitalization: Integrating TV and Internet?”, 7 June, *International Journal of Digital Television*, pp. 239-252.

Jenkins, H. (2006) “Convergence Culture – Where old and new media collide”. New York: University Press.

Jenkins, H., Ford S., and Green J. (2013) “Spreadable Media: Creating Value and Meaning in a Networked Culture (Postmillennial Pop)”. New York; University Press.

Jensen, K.B. (2014) “Audiences, Audiences Everywhere-Measured, Interpreted and Imagined”, in Patriarche G., Bilandzic, H., Jensen J.L. and Jurišić, J. (eds) *Audience Research Methodologies – Between Innovation and Consolidation.* Routledge: New York and London, pp. 227-237.

Jensen, K.B. (2011) “New media – old methods-Internet methodologies and the online/offline divide”, in Consalvo, M. and Ess, (eds), *The Blackwell Handbook of Internet Studies*, pp. 435-450. Malden, MA: Wiley-Blackwell.

Jensen, K.B. (2010) *Media Convergence – The three degrees of Network, Mass, and Interpersonal Communication*. London and New York: Routledge.

Jones, S., (1999) “Studying the Net Intricacies and Issues”, in *Doing Internet Research. Critical Issues and Methods for Examining the Net*, Jones, S. (ed). Sage Publications: 1-27.

Kale, A., Karandikar, A., Kolari, P., Java, A., Finin, T., and Joshi, A. (2007) “Modelling trust and influence in the blogosphere using link polarity”, in *Proceedings of the International Conference on Weblogs and Social Media* (ICWSM).

Kasapović M. (1999) “Die politische Entwicklung Kroatiens von 1990 – 1997” (The political development of Croatia 1990-1997), in *Der Jugoslawien-Krieg, Handbuch zu Vorgeschichte, Verlauf und Konsequenzen*, Melčić, Dunja (pub.), Westdeutscher Verlag, Opladen, Wiesbaden: 455-463.

Kepplinger, H.M. and Martin, V. (1986) “Functions of Mass Media in Interpersonal Communication”, *Communication*, (12), pp. 58-66.

Kennedy, H. (2012) “Perspective on Sentiment Analysis”, in *Journal of Broadcasting & Electronic Media*, Vol. 56, No. 4: 434-450.

Kern, E. (2012) “Game changer: Twitter rolls out expected restrictions to API use”, *Gigaom*, 16 August 2012 [Online], Available at: https://gigaom.com/2012/08/16/twitter-rolls-out-expected-restrictions-to-api-use/. (Accessed: 16th May 2016).

Kietzmann, J.H., Hermkens, K., McCarthy, I.P. and Silvestre, B.S. (2011) “Social media? Get serious! Understanding the functional building blocks of social media”, Business horizons, 54(3), pp.241-251.

Kim, P. and Sawhney, H. (2002), “A Machine-Like New Medium - Theoretical Examination of Interactive TV”, *Media, Culture & Society*, 24(2), pp.217-233.

KNIME (2014), “KNIME Analytics Platform”. Available at: https://www.knime.org/knime-analytics-platform (Accessed: 5th August 2014).

Konrad Adenauer Stiftung (2016) *Media and Democracy in South East Europe, Medienprogramm Südosteuropa*. Available at: http://www.kas.de/medien-europa/ (Accessed: 3rd Juni 2014).

Kreutzer R. (2014) *Notwendigkeit eines Change-Managements im Online-Zeitalter: Grundprinzipien zur erfolgreichen digitalen Transformation*. Springer Gabler. Berlin.

Lessig, L. (2005) “Commons on the Wires”, in *Creative Industries*, Hartley, J. (ed.). Oxford: Blackwell.

Lettaru, K., Wang, S., Cao G., Padmanabhan, A. and Shook, E. (2013) "Mapping the global Twitter heartbeat: The geography of Twitter", 6 May 2013, *First Monday* [Online], 18(5). Available at www.firstmonday.org
(Accessed 13th November 2013).

Levine, J. (2012) "The broadcastification of social media", 12 December 2012, Available at: http://www.niemanlab.org/2012/12/the-broadcast-ification-of-social-media/
(Accessed: 14th July 2014).

Library of Congress (2013) "Update on the Twitter Archive at the Library of Congress", January 2013. Available at:
https://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf
(Accessed: 23th February 2013).

Livingstone, S.M. (2005) "People living in the new media age: rethinking 'audiences' and 'users', paper presented at: *New Approaches to Research on The Social Implications of Emerging Technologies*". Oxford Institute, April 15th-16th.

Livingstone, S. (1999). "New Media, New Audience?" [online], London: LSE Research Online. Available at: http://eprints.lse.ac.uk/391/1/N-media& society1 (1).pdf
(Accessed: 15th February 2012).

Livingstone, S.M. (1998a) "Making Sense of Television – The Psychology of Audience Interpretation", *International Series in Social Psychology*, 2nd ed. Routledge.

Livingstone S. (1998b) "Relationships between media and audiences: prospects for audience reception studies" in Liebes, T. and Curran, J. (eds) *Media, Ritual and Identity: Essays in Honor of Elihu Katz*. London, Routledge, pp. 237-255.

Liu, B. (2010), "Sentiment Analysis and Subjectivity", in Indurkhya, N. and Damerau F. J. (eds) *Handbook of Natural Language Processing.* 2nd edn. CRC Press, pp. 627-666.

Liu, Y., Gummadi, K.P., Krishnamurthy, B. and Misolove, A. (2011) "Analyzing Facebook Privacy Settings: User expectations vs. reality", *Proceedings of the 2011 ACM SIGGOMM Conference on Internet Measurement Conference*. New York, pp. 61-70.

Liu, H., Lieberman H. and Selker, T. (2003) "A model of textual affect sensing using real-world knowledge", *Proceedings of Intelligent User Interfaces (IUI)*, January 12-15, 2003, Miami, Florida, pp. 125–132. Available at: http://larifari.org/_/writing/IUI2003-AffectSensing.pdf.
(Accessed: 15th May 2015).

Ljubešić, N., Fišer D. and Erjavec, T. (2014) "TweetCaT: a tool for building Twitter corpora of smaller languages", *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14)*. Reykjavik, pp. 2279-2283. Available at: http://nlp.ffzg.hr/data/publications/nljubesi/ljubesic14-tweetcat.pdf
(Accessed: 15th March 2014).

Lomborg, S. (2012) "Researching Communicative Practice: Web Archiving in Qualitative Social Media Research", *Journal of Technology in Human Services*, 30(3-4), pp. 219-231.

Lotz, A. (2007) *Television will be Revolutionized*. New York, University Press.

Lovejoy, K. and Saxton, G.D. (2012) "Information, community, and action: How non-profit organizations use social media", *Journal of Computer-Mediated Communication*, *17*(3), pp.337-353.

Lula, P. and Wojcik, K. (2011) "Sentiment Analysis of Consumer Opinions written in Polish", *Economics and Management*, 16, p.1286-1291. Available at: http://internet.ktu.lt/lt/mokslas/zurnalai/ekovad/16/1822-6515-2011-1286.pdf. (Accessed: 15th June 2014).

Manyika, J., Chui, M., Brad Brown, B., et al. (2011) "Big data: The next frontier for innovation, competition, and productivity", *McKinsey Global Institute*. Available at: http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation
(Accessed: 18th October 2015).

Malović, Stjepan (2007) *Vjerodostojnost novena* (Newspaper Credibility). Zagreb: ICEJ i Sveučilišta knjižara d.o.o.

Manovich, L. (2012) "Trending: The Promises and the Challenges of Big Social Data", Gold M. K. (ed.) *Debates in the Digital Humanities*. *Minneapolis*, *MN*: The *University* of *Minnesota* Press, pp.460-475.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think,* Houghton Mifflin Harcourt.

McAfee, A. and Brynjolfsson, E. (2012) "Big Data: The Management Revolution", *Harvard Business Review*, October 2012, pp. 61-68.

McQuail, D. (2000) *McQuail's Mass Communication Theory*, 4th edition. London: Sage Publications.

McQuail, D. (1997) *Audience Analysis*. USA: Sage Publications, Inc.

Michelson, M. and Macskassy, S.A. (2010) "'Discovering users' topics of interest on twitter: a first look", In *Proceedings of the fourth workshop on Analytics for noisy unstructured text data* (AND '10). ACM, New York, NY, USA, pp. 73-80. Available at: http://www.research.rutgers.edu/~sofmac/paper/and2010/michelson-and2010.pdf (Accessed: 20th February 2014).

Milavsky, J. R. (1992) "How good is the A. C. Nielsen People-Meter System? A Review of the Report by the Committee on Nationwide Television Audience Measurement", *Public Opinion Quarterly*, Oxford University Press, 56(1), pp. 102-115.

Mitra, A., and Cohen, E., (1999) "Analyzing the Web Directions and Challenges", Jones, S. (ed.) *Doing Internet Research*. *Critical Issues and Methods for Examining the Net.* Sage Publications, pp. 1-27.

Morinaga, S., Yamanishi K., Tateishi, K. and T. Fukushima (2002) "Mining product reputations on the web", *Proceedings of the ACM SIGKDD, Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 341–349.

Napoli P. (2012) "Program Value in the Evolving Television Audience Marketplace", Time Warner Research Program on Digital Communications.

Napoli P. (2011) “Ratings and Audience Measurement”, Nightingale, V. (ed.) *The Handbook of Media Audiences,* pp. 286-301.Wiley-Blackwell. Available at: http://www.citi.columbia.edu/B8210/read24/ch14.pdf
(Last accessed: 15th February 2013).

Nasukawa, T. and Yi, J. (2003) “Sentiment analysis: Capturing favorability using natural language processing”, *Proceedings of the Conference on Knowledge Capture (K-CAP).*

Neufeld, J. (2011) “iPads, Mobiles, & Social Media: The Perfect TV Companions”, IBC (International Broadcasting Convention).

Newmann, C. (2012) “Sentiment140 resource report: collecting and analyzing Twitter sentiment”, *The News and Social Technologies*, Available at: jacdigital.com
(Accessed: 15th May 2013).

Nielsen (2016) “Nielsen to launch social content ratings with measurement across twitter and Facebook”, 20th January. Available at: http://www.nielsen.com/us/en/press-room/2016/nielsen-to-launch-social-content-ratings-with-measurement-across-twitter-and-facebook.html
(Accessed: 4th March 2016).

Nielsen (2015) “Screen Wars – The Battle for Eye Space in a TV-Everywhere World” (March 2016) [Online]. Available at: http://www.nielsen.com/content/dam/corporate/Italy/reports/2015/Nielsen%20Global%20Digital%20Landscape%20Report%20March%202015.pdf
(Accessed: 4th July 2015).

Nielsen (2014) “Nielsen launches demographics for Nielsen Twitter TV Ratings”, 19th May. Available at: http://www.nielsen.com/tz/en/press-room/2014/nielsen-launches-demographics-for-nielsen-twitter-tv-ratings.html
(Accessed: 4th March 2015).

Nielsen (2013a) “Nielsen launches Nielsen TV ratings”, 7th October. Available at: http://www.nielsen.com/us/en/press-room/2013/nielsen-launches-nielsen-twitter-tv-ratings.html
(Accessed: 12th December 2013).

Nielsen (2013b) “New study confirms correlation between Twitter and TV Ratings”, 7th October 2013. Available at: http://www.nielsen.com/ca/en/press-room/2013/new-study-confirms-correlation-between-twitter-and-tv-ratings.html
(Accessed: 12th December 2013).

Nielsen (2012a) “Nielsen and Twitter establish Social TV Ratings”, 7th December 2023. Available at: http://www.nielsen.com/us/en/press-room/2012/nielsen-and-twitter-

Nielsen (2012b), “State of the Media | The Social Media Report 2012”.

Nielsen (2013). Available at: http://www.nielsen.com/hr/hr.html
(Accessed: 1 January 2013).

Noguera, J.M., Villi, M., Nyiro, N., De Blasio, E. and Bourdaa, M. (2014) “The Role of the Media Industry when Participation is a Product”, in Carpentier, N., Schroeder, K., and Hallett, L. (eds) *Audience Transformations: Shifting Audience Positions in Late Modernity*. Routledge: UK, pp. 172-190.

Ounis I., Macdonald C., and Soboroff I. (2008) “On the TREC Blog Track”, *Proceedings of the International Conference on Weblogs and Social Media* (ICWSM).

Pang B., and Lee L. (2008) “Opinion mining and sentiment analyses”, *Foundations and Trends in Information Retrieval*. 2(1-2), pp. 1-135. Available at: http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf.
(Accessed: 15th March 2013).

Pang, B., Lee, L. and Vaithyanathan, S. (2002) “Thumbs up? Sentiment classification using machine learning techniques”, *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.79–86. Available at: http://www.cs.cornell.edu/home/llee/papers/sentiment.pdf
(Accessed: 18th March 2013).

Panian, Ž. (2010) “Text Analytics as a Form of Knowledge Mining”, Grigoriu, M. Mladenov, V., Bulucea, A. C., Martin, O., and Mastorakis, N. (eds) *Proceedings of the 4th European Computing Conference (ECC ‘10)*, Bucharest, Romania. WSEAS Press, Athens, pp. 112-117.

Pasquali, F. (2012) “Networked Audiences and Small Scale Groups Belongings: Viewing, Sharing and Archiving TV Content in the Italian Social Media Scenario”, Observatorio (*OBS) Special Issue Networked belonging and networks of Belonging,* pp.181-196. Available at: http://obs.obercom.pt/index.php/obs/article/viewFile/583/524
(Accessed: 14th July 2014).

Patriarche, G., Bilandzic, H., Jensen, J.L. and, Jurisic, J. (2014) “Introduction”, in Patriarche, G., Bilandzic, H., Jensen, J.L. and, Jurisic, J. (eds) *Audience Research Methodologies – Between Innovation and Consolidation*. Routledge: New York and London, pp.196-211.

Penn State University (2015) “Information by Language”. Available at: http://symbolcodes.tlt.psu.edu/bylanguage/serbocroatian.html (Accessed: 19th February 2015).

Perez, S. (2014) “Nielsen Twitter TV Ratings now identify the age and gender of those tweeting about TV or just lurking”, 19th May, *techcrunch* [Online]. Available at: https://techcrunch.com/2014/05/19/nielsens-twitter-tv-ratings-now-identify-the-age-and-gender-of-those-tweeting-about-tv-or-just-lurking/ (Accessed 4th August 2016).

Pew Research Center (2013) “Social Networking Fact Sheet”, 27th December. Available at: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ (Accessed: 14th April 2014).

Peruško, Z. (2011) “Assessment of media development in Croatia: based on UNESCO’s media development indicators”. Available at: http://www.unesco.org/new/fileadmin/mulimedia/HQ/CI/CI/pdf/IPDC/croatia_report_final.pdf (Accessed: 19th March 2016).

Phelan, O., McCarthy, K., Smyth, B.(2009)”Using Twitter to recommend real-time topical news”, *Proceedings of the third ACM conference on Recommender systems*, Paper presented at the 3rd ACM Conference on Recommender Systems (RecSys 2009), New York City, NY, USA, 22-25 Available at: http://hdl.handle.net/10197/1893 (Accessed: 15th November 2014).

Plaesu, A. and Dalu, A. (2008) “The social and interactional dimensions of Internet utilization”, *Recent Advances in Education and Educational Technology*, WSEAS Press, pp. 198-202.

Plunkett, J. (2012) “Don't break stories on Twitter, BBC journalists told”, 8th February, *The Guardian* [Online]. Available at: https://www.theguardian.com/media/2012/feb/08/twitter-bbc-journalists (Accessed: 24th April 2013).

Prescott Thomas, J. (2007) “Media Management Manual’. A Handbook for television and radio parishioners in countries-in-transition”. Unesco House. Available at: http://unesdoc.unesco.org/images/0018/001879/187966e.pdf (Accessed: 12th February 2016).

Prior, M. (2007) *Post-Broadcast Democracy: How Media Choice Increases Inequality Political Involvement and Polarizes Elections*. Cambridge, NY: Cambridge University Press.

Procter, R., Voss, A., and Lvov, I. (2015) “Audience research and social media data: Opportunities and challenges”, *Participation Journal of Audience & Reception Studies,* 12(1), pp. 470-493. Available at: http://participations.org/Volume%2012/Issue%201/29.pdf (Accessed: 12th June 2015).

Proulx, M. and Shepatin, S. (2012) *Social TV: How Marketers can Reach and Engage Audiences by Connecting Television on the Web, Social Media and Mobile*, Hoboken, New Jersey: Wiley.

Ravindranath, M. (2014) “Twitter grants selected researchers access to its public database”, 26th October. Washington Post [Online], Available at: https://www.washingtonpost.com/business/on-it/twitter-grants-select-researchers-access-to-its-public-database/2014/10/26/48a1f532-5931-11e4-b812-38518ae74c67_story.html
(Accessed: 23rd November 2014).

Reitze, H., Ridder, C.M. (eds) (2011) *Massenkommunikation VIII - eine Langzeitstudie zur Mediennutzung und Medienbewertung 1964-2010*. Schriftenreihe *Media Perspektiven*, Band 21. Nomos Verlagsgesellschaft: Baden Baden.

Reinold, A. (1994) “Patterns of Individual TV Usage”, *Worldwide Electronic and Broadcast Audience Research Symposium*, PARIS (France) 1st - 4th May 1994, Amsterdam, ESOMAR, pp. 67-78.

Riff, D., Lacy, S. and Fico, F. (2014) *Analyzing media messages: Using quantitative content analysis in research*. Routledge.

Rogers, R. (2013a) *Digital Methods*. The Mit Press: Cambridge Massachusetts.

Rogers, R. (2013b) “Debanalizing Twitter: The Transformation of an Object of Study”, *WebSci ’13 Proceedings of the 5th Annual ACM Web Science Conference*, pp. 356-365. Available at: http://govcom.org/publications/full_list/rogers_debanalizingTwitter_websci13.pdf
(Accessed: 15th September 2014).

Rosen, J. (2006) “The People Formerly Known as the Audience”, 27 June, *Press Think* [Online], 27 June 2006. Available at:

http://www.archive.pressthink.org/2006/06/27/ppl_frmr.html
(Accessed: 21st January 2014).

RTL Group (2014) "Television 2014 – International Key Facts", 21th edn, IP International Marketing committee (ed). Luxembourg.

RTL Group (2013) "Television 2013 – International Key Facts" 20th edn, IP International Marketing committee (ed). Luxembourg.

RTL Group (2012) "Television 2012 – International Key Facts" 19th edn, IP International Marketing committee (ed). Luxembourg.

RTL Group (2008) "Television 2008 – International Key Facts" 15th edn, IP International Marketing committee (ed). Luxembourg.

Runje, D. (2013) "Hrvati preferiraju Facebook, ali popularnost Twitter raste", 15th February, *Poslovni dnevnik* [Online]. Available at: http://www.poslovni.hr/marketing-i-mediji/hrvati-preferiraju-facebook-ali-popularnost-twittera-snazno-raste-230772
(Accessed:15th February 2013).

Sachs, L. (2002) *Angewandte Statistik,* 9th edn.Springer-Verlag, Berlin, Heidelberg New York.

Scola, N. (2015) "Library of Congress" Twitter archive is a huge #FAIL, 7 November, *Poltico* [Online]. Available at: http://www.politico.com/story/2015/07/library-of-congress-twitter-archive-119698.html
(Accessed: 23rd April 2016).

Sentiment140 (2013) [opinion mining tool]. Available at:
http://www.sentiment140.com/
(Accessed: 26th February 2013).

Silverstone, R. (2003) *Television and Everyday Life*. Taylor & Francis e-Library [Online]. 2nd edn. 1994, Routledge: London.

Socialmention (2013) [opinion mining tool] Available at: http://socialmention.com.
(Accessed: 26th February 2013).

Somasundaran, S., Wilson T., Wiebe, J. and Stoyanov V. (2007), "QA with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news", *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*. Available at:
http://people.cs.pitt.edu/~swapna/papers/socMedia.v10.pdf
(Accessed: 4th April 2013).

Splichal, S. (2007) "Does History Matter? Grasping the Idea of Public Service at its Roots", pp. 237-256, in Bardoel, J. and Lowe G.F. *From Public Service Broadcasting to Public Service Media.* Nordicom.

Statista (2016a) "Global Social Networks ranked by Number of Users". Available at:
http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/
(Accessed: 1st April 2015).

Statista (2016b) “Spending on Advertising in Croatia”. Available at: http://www.statista.com/statistics/491737/advertising-spend-by-medium-croatia/ (Accessed: 10th March 2016).

Statista (2016c) “The last time you were using the internet while watching TV, which of the following devices did you use for going online?” Available at: http://www.statista.com/statistics/369809/second-screen-device-usage-croatia/ (Accessed: 10th March 2016).

Statista (2014) “Number of TV Households in the United States”. Available at: http://www.statista.com/statistics/243789/number-of-tv-households-in-the-us/ (Accessed: 10th March 2014).

Strapparava, C. and Valitutti, A. (2004) “WordNet-Affect: An affective extension of WordNet”, in *Proceedings of LREC*, pp.1083–1086. Available at: http://wndomains.fbk.eu/publications/lrec2004.pdf (Accessed: 15th June 2013).

Sudweeks, F and Simoff S. J. (1999) ´Complementary Explorative Data Analysis the Reconciliation of Quantitative and Qualitative Principles, pp. 29-55, Jones, S. (ed.) *Doing Internet Research. Critical Issues and Methods for Examining the Net.* Sage Publications. Available at: http://www.it.murdoch.edu.au/~sudweeks/papers/ceda.pdf (Accessed: 14th October 2014).

Sweetscape [text editor software] http://www.sweetscape.com/010editor/

Talbot, D. (2011) "*A Social-Media Decoder*", 18th October, MIT Technology Review [Online]. Available at: http://www.technologyreview.com/featuredstory/425787/a-social-media-decoder/
(Accessed: 20th May 2016).

Tateishi, K., Ishiguro, Y and Fukushima, T. (2001) "Opinion information retrieval from the Internet", *Information Processing Society of Japan (IPSJ)* SIG Notes, pp.75–82.

Tong, R. M. (2001) "An operational system for detecting and tracking opinions in on-line discussion", *Proceedings of the Workshop on Operational Text Classification* (OTC).

Topsy [opinion mining tool] Available at http://analytics.topsy.com/ (Accessed: 26th February 2013).

Tavakoli M. and Cave M. (1996) "Modelling Television Viewing Patterns", *Journal of Advertising*, 25(4), pp. 71-86.

Thelwall, M., Buckley, K. and Paltoglou, G. (2011), "Sentiment in Twitter events", *Journal of the American Society for Information Science and Technology*, 62(2), pp.406-418.

Thompson, W., Mo, H., Pacheco, J. and Carroll, R. (no date) "Developing Executable Phenotype Algorithms Using the KNIME Analytics Platform", Available at: http://informatics.mayo.edu/phema/images/4/40/KNIME_Tutorial_v1.pdf
(Accessed: 15th February 2015).

Torrez-Riley, J. (2011), "The Social TV Phenomenon: New Technologies Look to Enhance Television`s Role as an Enabler of Social Interaction". Elon, NC: Elon University. Available at: http://www.torrezriley.com/projects/researchpaper/SocialTVpaper.pdf (Accessed: 12th May 2014).

Turney, P. (2002), "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 417–424. Available at: http://www.aclweb.org/anthology/P02-1053.pdf (Accessed: 15th June 2013).

Tweetfeel [Opinion Mining Tool]. Available at: www.tweetfeel.com. (Accessed: 26th February 2013).

Tweetreach [Opinion Mining Tool]. Available at: https://tweetreach.com/ (Accessed: 26th February 2013).

Twitter (2016) "The Twitter user agreement". Available at: https://twitter.com/tos?lang=en#twitterrights. (Accessed: 20th May 2016).

Twitter (2014) "Twitter #datagrants selections", 17th April. Available at: https://blog.twitter.com/2014/twitter-datagrants-selections (Accessed: 17th April 2014).

Twitter (2014) “Introducing Twitter Data Grants”, 15th February. Available at: https://blog.twitter.com/2014/introducing-twitter-data-grants.
(Accessed: 15th February 2014).

Twitter Developers (2011) “The Search API”, Twitter Developers, Available at: https://dev.twitter.com/rest/public/search
(Accessed: 1st August 2013).

Twitter (2011) “Twitter, TV and You”, 26 October, Available at: https://blog.twitter.com/2011/twitter-tv-and-you
(Accessed: 13th November 2013).

Van Eimeren, B. and Frees, B. (2009) „Der Internetnutzer 2009 - multimedial und total vernetzt? Ergebnisse der ARD/ZDF-Onlinestudie 2009“, *Media Perspektiven* (7), pp. 334-348.

Vinayagamoorthy, V. and Kramskoy, J. (2013) “An Experiment in Social TV for Automating Content Sharing via Social Networks”. *BBC Research & Development WHP 243*. Available at: http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP243.pdf
(Accessed: 20th May 2016).

Vincente-Marino, M. (2014) “Audience Research Facing the Challenges of Transforming Audiences”, in Patriarche, G., Bilandzic, H., Jensen, J.L. and, Jurisic, J. (eds) *Audience Research Methodologies – Between Innovation and Consolidation*. Routledge: New York and London, pp.196-211.

Wakamiya, S., Lee, R. and Sumiya, K. (2011), "Towards better TV viewing rates: exploiting crowd's media life logs over Twitter for TV rating", *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*, ACM, (p. 39).

Webster, T. (2010) "Twitter Usage in America 2010. The Edison Research/Arbitron Internet and Multimedia Study", *Edison Research*. Available at: http://www.edisonresearch.com/twitter_usage_in_america_2010_1/ (Accessed: 15th February 2013).

Webster J.G., Phalen P.F. and Lichty W.L. (2007) *Ratings Analysis – The Theory and Practice of Audience Research*, 3rd edn. Lawrence Erlbaum Associates. Mahwah, New Jersey London.

Wiebe, J., Breck, E., Buckley, C., Cardie, C., Davis, P., Fraser, B., Litman, D., Pierce, D., Riloff, E., Wilson, T., Day, D., and Maybury, M. (2003) "Recognizing and organizing opinions expressed in the world press", *Proceedings of the AAAI Spring Symposium on New Directions in Question Answering*. Available at: https://www.cs.cornell.edu/home/cardie/papers/aaai-ss-overall-03.pdf. (Accessed: 14th December 2013).

Williams, F. L. (1972) "Analysis of Life Styles for Study in Consumer Behavior", *SV - Proceedings of the Third Annual Conference of the Association for Consumer Research*, Venkatesan; M. (ed.), Chicago, IL: Association for Consumer Research, pp. 490-501. Available at: http://acrwebsite.org/volumes/12023/volumes/sv02/SV-02 (Accessed: 15th April 2014).

Wilson T., Wiebe J. and Hoffmann P. (2005) "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", *Proceedings of the Human Language Technologies Conference/ Conference on Empirical Methods in Natural Language Processing*: 347–354. Available at: http://people.cs.pitt.edu/~twilson/pubs/hltemnlp05.pdf (Accessed: 14th March 2013).

Wikipedia Available at: https://www.wikipedia.org/

Wimmer, R.D. and Dominick J.R. (2014) *Mass Media Research: An Introduction*, 10th edn. Wadsworth, Cengage Learning (USA).

Wohn, D.Y. and Na, E.-K. (2011) "Tweeting about TV: Sharing television viewing experiences via social media message streams", 7th March, *First Monday* [Online], 16(3). Available at: http://firstmonday.org/article/view/3368/2779 (Accessed: 4 October 2012).

Wu, X., Zhu, X., Wu, G.Q. and Ding, W. (2014), "Data mining with big data", IEEE *transactions on knowledge and data engineering*, 26(1), pp.97-107. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.419.9182&rep=rep1&type=pdf (Accessed: 15th July 2015).

Yin, S. (2011) "Twitter Cracks Down on Another Third-Party App, Twapperkeeper", 23 February, PC [Online]. Available at: http://www.pcmag.com/article2/0,2817,2380784,00.asp (Accessed 5th November 2013).

Yourtwapperkeeper (2013) Archive your Social Media. Available at: https://github.com/540co/yourTwapperKeeper (Accessed: 27th March 2013).

Yu, H. and Hatzivassiloglou, V. (2003) “Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences”, *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Available at: http://www.aclweb.org/anthology/W03-1017.pdf (Accessed: 15th March 2013).

Zabin J. and Jefferies A. (2008) “Social media monitoring and analysis: Generating consumer insights from online conversation”, *Aberdeen Group Benchmark Report*. Available at: https://robertoigarza.files.wordpress.com/2008/10/rep-social-media-monitoring-and-analysis-aberdeen-group-2008.pdf (Accessed: 12th February 2014).

Zellig, H. (1954) “Distributional Structure”, *Word* 10(23), pp. 146–162.

Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H. and Li, X., 2011, April. Comparing Twitter and Traditional Media Using Topic Models. In European Conference on Information Retrieval (pp. 338-349). Springer Berlin Heidelberg.

Züll, B., Boras, D. and Mikelić Preradović, N. (2013) “The Digital Change of Television in a Social Media Environment: The case of the BBC”, *Proceedings of the 7th European Computing Conference (ECC’13)*, pp. 276-281. WSEAS Press. Available at: http://www.wseas.us/e-library/conferences/2013/Dubrovnik/ECC/ECC-44.pdf (Accessed: 5th August 2016).

Züll, B. and Mikelić Preradović, N. (2013) “Methods and usage of sentiment analysis in the context of the TV industry”, *Proceedings of the 7th European Computing Conference (ECC’13*), pp.124-129. WSEAS Press: 124-129. Available at: http://www.wseas.us/e-library/conferences/2013/Dubrovnik/ECC/ECC-19.pdf (Accessed: 5th August 2016).

Züll, B. (2013) “Protecting Forgetting and the ‘Right to be forgotten’ in the Context of Social Media”, *Proceedings of the 7th European Computing Conference (ECC’13)*, pp. 210-215. WSEAS Press. Available at: http://www.wseas.us/e-library/conferences/2013/Dubrovnik/ECC/ECC-33.pdf
(Accessed: 5th August 2016).

Züll, B. (2013) “The Dilemma of Internet Freedom”, *Proceedings of the 7th European Computing Conference (ECC’13)*, pp. 240-245.WSEAS Press. Available at: http://www.wseas.us/e-library/conferences/2013/Dubrovnik/ECC/ECC-38.pdf
(Accessed: 5th August 2016).

Züll, B. (2014) “The Vision of Global Internet Freedom”, *International Journal of Computers and Communication*, pp. 120-127, ISSN: 2074-1294, Volume 8 Available at: http://www.naun.org/main/UPress/cc/2014/a062012-119.pdf
(Accessed: 5th August 2016).

TABLES

2.1 Classification of social media

2.2 Country data of Croatia

2.3 Digital television in Croatia, 2013.

2.4 Television landscape in Croatia

2.5 Domestic public television landscape in 2013

2.6 Domestic private television landscape in 2013

2.7 McQuail`s three audience research traditions

3.1 Panel size

3.2 Market shares and TV ratings of the most popular television programs in Croatia 2013

3.3 Age cluster of Croatian television audience

3.4 Life style of panel households

3.5 Social classes of panel members

4.1 Twitter data collection processing framework

5.1 Selection of opinion mining tools used for this research

5.2 Definition of search terms

5.3 Distribution of tweets following language code

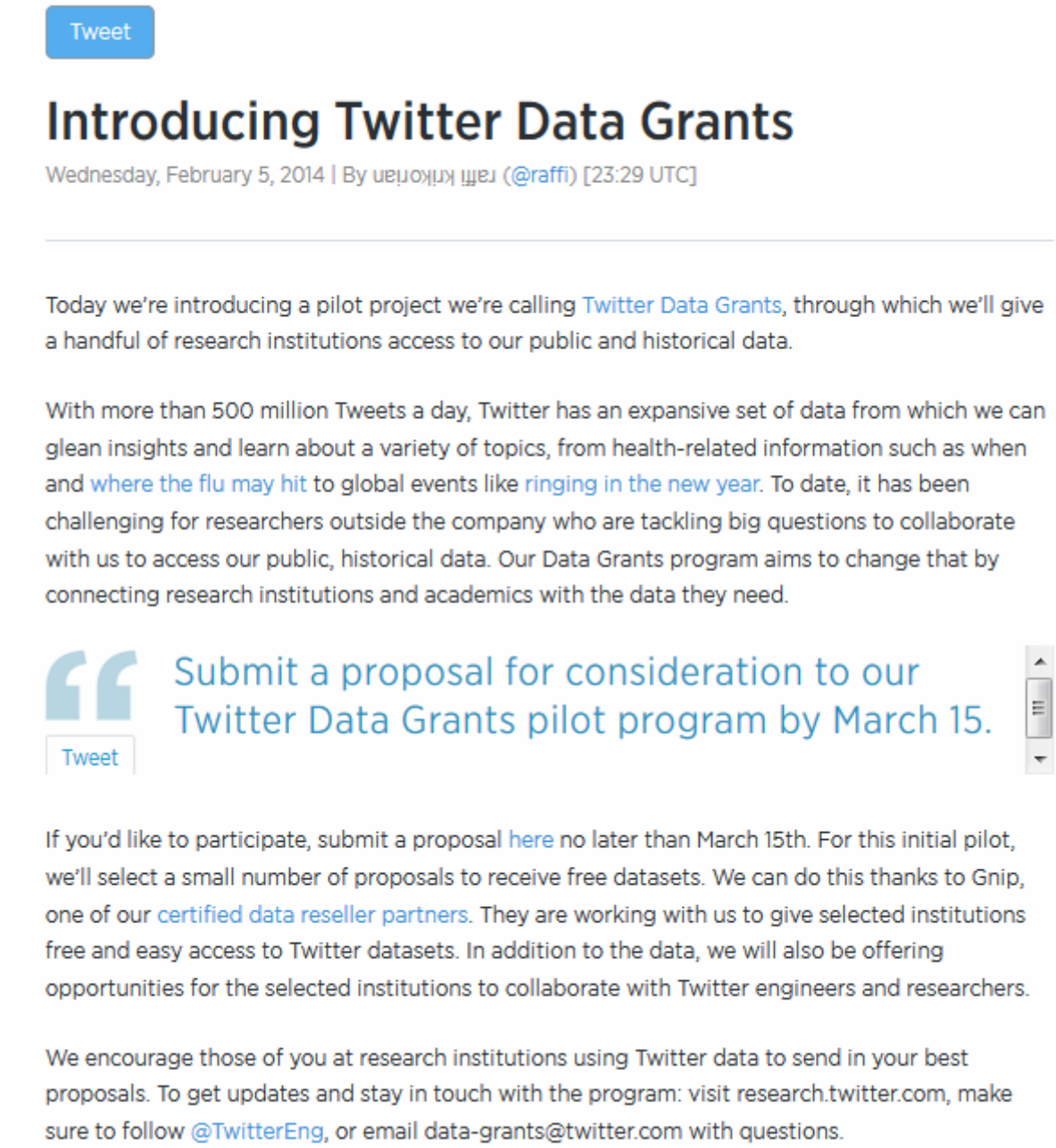
5.4 Top 30 - Tf-idf ranking of television-related tweets

FIGURES

2.1 Development of television and social media
2.2 Internet penetration in Europe November 2015
2.3 Media usage in Croatia
2.4 Levels of television viewing time of individuals in Croatia (aged 4+)
2.5 Social television and second-screen usage
2.6 Leading social networks worldwide (December 2014), ranked by number of active users in millions
2.7 What people share on social networks
2.8 Spending on advertising years 2007 to 2015 in Croatia in million Kuna
2.9 Audience shares of main channels in 2011
2.10 Audience shares of main channels in 2012
2.11 Audience shares of main channels in 2013
3.1 People-meter configuration
3.2 Map of Croatia, regions 1-6
3.3 Television ratings
3.4 Market shares
3.5 Viewing patterns in Croatia 2013
3.6 Sample-feature Representation
4.1 Ranking of social media site in Croatia
4.2 Usage number of Twitter in Croatia
4.3 Quote
5.1 Opinion mining using Sentiment140: Example *Oscars*.
5.2 “Positive” statement towards #*oscars*
5.3 “Negative” statement towards #*oscars*
5.4 “Negative statement towards #americandad
5.5 Opinion mining with feelings *oscars*
5.6 Opinion mining with feelings *american dad*
5.7 Opinion mining with feelings *Who wants to be a millionaire*
5.8 Searching the opinion with Tweettone. Example: *Oscars*
5.9 Searching the opinion with Tweettone. Example: *Who wants to be a millionaire*
6.1 Methodological framework

APPENDIX 1

Figure 1: Introducing Twitter Data Grants (2014)



APPENDIX 2

Application for Twitter data grant (March 2014)

Proposal name:

Analysis of Twitter and the public sphere in Croatia, Serbia and Bosnia and Herzegovina

Proposal description:

The aim of our research is to analyse the Twitter data in Croatia, Serbia and Bosnia and Herzegovina.

We search for conversation on Twitter in context of relevant political, media and other topics of public interest to find out if and how Twitter is used by the Croatian, Serbian, and Bosnian public.

We analyse if Twitter and how it is used as communication instrument in our societies.

We plan to do research on the following topics:

- War related trials in Den Haag
- European Elections
- Entrance in the European Union of Croatia
- Strikes and protests in Croatia/Serbia/Bosnia on certain public topics such as health personal strike in Croatia
- Political and economic trials with high public interest such as Sanader (ex. Prime Minister) trial
- Media Events with high public interest

As Croatians, Serbians and Bosnians mainly tweet in Croatian/Serbian/Bosnian language the main task is to identify the Croatian/Serbian/Bosnian Twitter corpus and then by sampling to identify the relevant topics. By Twitter corpus we mean all Tweets written in Croatian/Serbian/Bosnian language (identifier = language) and Tweets written by Croatian/Serbians/Bosnian in English (identifier = coordinates).

Funding

Funding is provided by Poslovna inteligencija, Croatian company specialized in Big Data analytics. They will provide us with technical resources / expertise and infrastructure.

Social data experience

- Processing and analysis of big data (social media)
- Opinion Mining/Sentiment analysis
- Language recognition library/ algorithm

Technical experience:

We have experience in various big data technologies:

- Hadoop and HBase
- MapReduce, Hive and Pig programming
- Storm for real time data processing
- Collecting, storing and processing JSON data through API
- KNIME for sentiment analysis

Filters advanced:

We need as filter the language or territorial filter. That means the creation of the Twitter the corpus for our region for a defined time period (01/01/2013 - 15/03/2014). With Twitter corpus we mean all Tweets written in Croatian/Serbian/Bosnian language (identifier = language) and Tweets written by Croats, Serbians and Bosnians in English (identifier = coordinates). We do not ask for specific keywords or hashtags as filter, because Twitter usage in Croatia, Serbia and Bosnia and Herzegovina is still in an early adapter phase and hashtags and keywords are not so much used in our region, so we could not get a significant database to do analysis only by keyword search.

Team:

Stjepan Malović, Ph.D. (Department of Information & Communication Sciences)

Damir Boras, Ph.D. (Dean of the Faculty of Philosophy)

bacc. oec. Hrvoje Gabelica (Faculty of economics - business informatics)

bacc. oec. Živko Krstić (Faculty of economics - business informatics)

Beatrice Züll (Department of Information & Communication Sciences)

APPENDIX 3

January 2013

Update on the Twitter Archive At the Library of Congress

In April, 2010, the Library of Congress and Twitter signed an agreement providing the Library the public tweets from the company's inception through the date of the agreement, an archive of tweets from 2006 through April, 2010. Additionally, the Library and Twitter agreed that Twitter would provide all public tweets on an ongoing basis under the same terms. The Library's first objectives were to acquire and preserve the 2006-10 archive; to establish a secure, sustainable process for receiving and preserving a daily, ongoing stream of tweets through the present day; and to create a structure for organizing the entire archive by date. This month, all those objectives will be completed. To date, the Library has an archive of approximately 170 billion tweets.

The Library's focus now is on confronting and working around the technology challenges to making the archive accessible to researchers and policymakers in a comprehensive, useful way. It is clear that technology to allow for scholarship access to large data sets is lagging behind technology for creating and distributing such data. Even the private sector has not yet implemented cost-effective commercial solutions because of the complexity and resource requirements of such a task. The Library is now pursuing partnerships with the private sector to allow some limited access capability in our reading rooms. These efforts are ongoing and a priority for the Library.

This document summarizes the Library's work to date and outlines present-day progress and challenges.

Why the Twitter Collection is Important to the Nation's Library

Twitter is a new kind of collection for the Library of Congress, but an important one to its mission of serving both Congress and the public. As society turns to social media as a primary method of communication and creative expression, social media is supplementing and in some cases supplanting letters, journals, serial publications and other sources routinely collected by research libraries.

Archiving and preserving outlets such as Twitter will enable future researchers access to a fuller picture of today's cultural norms, dialogue, trends and events to inform scholarship, the legislative process, new works of authorship, education and other purposes.

The Library of Congress Agreement with Twitter

The Library's agreement with Twitter announced April 14, 2010 provided that:

- Twitter would donate a collection consisting of all public tweets from the Twitter service from its inception to the date of the agreement, an archive of 21 billion tweets that occurred between 2006 and 2010.
- Any additional materials Twitter provides to the Library would be governed by the terms of the agreement unless both parties agree to different terms in advance of receiving such additional materials.
- The Library could make available any portion of the collection six months after it was originally posted on Twitter to "bona fide" researchers.
- A researcher must sign a "notification" prohibiting commercial use and redistribution of the collection.
- The Library cannot provide a substantial portion of the collection on its web site in a form that can be easily downloaded.

Transfer of Data to the Library

In December, 2010, Twitter named a Colorado-based company, Gnip, as the delivery agent for moving data to the Library.

Shortly thereafter, the Library and Gnip began to agree on specifications and processes for the transfer of files – "current" tweets - on an ongoing basis.

In February 2011, transfer of "current" tweets was initiated and began with tweets from December 2010.

On February 28, 2012, the Library received the 2006-2010 archive through Gnip in three compressed files totaling 2.3 terabytes. When uncompressed the files total 20 terabytes. The files contained approximately 21 billion tweets, each with more than 50 accompanying metadata fields, such as place and description.

As of December 1, 2012, the Library has received more than 150 billion additional tweets and corresponding metadata, for a total including the 2006-2010 archive of approximately 170 billion tweets totaling 133.2 terabytes for two compressed copies.

Building a Stable, Sustainable Archive

The Library's first and most fundamental activities included developing a stable and sustainable way to acquire, preserve and organize the Twitter collection.

Although the Library regularly acquires digital content, the Twitter stream is the first collection coming into the Library in a continuous stream. The Library leveraged the technical infrastructure and workflow established for other digital content in the transfer of Twitter data.

The Library runs a fully automated process for taking in these new files. Gnip, the designated delivery agent for Twitter, receives tweets in a single real-time stream from

Twitter. Gnip organizes the stream of tweets into hour-long segments and uploads these files to a secure server throughout the day for retrieval by the Library.

When a new file is available, the Library downloads the file to a temporary server space, checks the materials for completeness and transfer corruption, captures statistics about the number of tweets in each file, copies the file to tape, and deletes the file from the temporary server space.

The technical infrastructure for the Library's Twitter archive follows the same general practices for monitoring and managing other digital collection data at the Library. Tape archives are the Library's standard for preservation and long-term storage. Files are copied to two tape archives in geographically different locations as a preservation and security measure.

The volume of tweets the Library receives each day has grown from 140 million beginning in February, 2011 to nearly half a billion tweets each day as of October, 2012.

The Library is processing data from the original 2006-2010 archive and organizing the material into hourly files. This operation is necessary so the entire archive from 2006 moving forward is organized the same – by time and in hourly files. This process will be completed in January 2013.

Toward Providing Collection Research Access

As with any collection, the Twitter archive must be processed and organized in a way that makes it useable. It is not uncommon for the Library to spend months or in some cases years sorting a large acquisition to inventory, organize and catalogue the information and materials so they are accessible by researchers.

The Library has extensive expertise in managing acquisition of and access to large-volume digital collections. For example, since 2000, the Library has been collecting web sites documenting government information and policy events. Today, that archive is more than 300 terabytes in size, and represents tens of thousands of web sites. Because there was a community of cultural heritage institutions and national libraries committed to collecting web sites, standards and tools have been developed collaboratively for capturing and providing access to these materials.

The Twitter Archive represents a new type of collection. The Twitter collection is not only very large, it also is expanding daily, and at a rapidly increasing velocity. The variety of tweets is also high, considering distinctions between original tweets, re-tweets using the Twitter software, re-tweets that are manually designated as such, tweets with embedded links or pictures and other varieties.

The Library has received approximately 400 inquiries from researchers all over the world since the announcement that it would accept the Twitter archive. Some broad topics of interest expressed by researchers thus far run from patterns in the rise of citizen

journalism and interest in elected officials' communications to tracking vaccination rates and predicting stock market activity. Many inquiries would inform research with policy and regulatory usefulness, such as tracking flu pandemic, citizen responses to candidates' stances on various issues and tracking public access to court systems. The nature of the queries also varies. For example, requests range from searching for a specific hashtag term to requesting a statistically valid sample of the entire stream.

What kind of information might researchers learn from the Twitter archive?

Some examples of the types of requests the Library has received indicate how researchers might use this archive to inform future scholarship:

* A master's student is interested in understanding the role of citizens in disruptive events. The student is focusing on real-time micro-blogging of terrorist attacks. The questions focus on the timeliness and accuracy of tweets during specified events.

* A post-doctoral researcher is looking at the language used to spread information about charities' activities and solicitations via social media during and immediately following natural disasters. The questions focus on audience targets and effectiveness.

The Library has not yet provided researchers access to the archive. Currently, executing a single search of just the fixed 2006-2010 archive on the Library's systems could take 24 hours. This is an inadequate situation in which to begin offering access to researchers, as it so severely limits the number of possible searches.

The Library has assessed existing software and hardware solutions that divide and simultaneously search large data sets to reduce search time, so-called "distributed and parallel computing". To achieve a significant reduction of search time, however, would require an extensive infrastructure of hundreds if not thousands of servers. This is cost-prohibitive and impractical for a public institution.

Some private companies offer access to historic tweets but they are not the free, indexed and searchable access that would be of most value to legislative researchers and scholars.

It is clear that technology to allow for scholarship access to large data sets is not nearly as advanced as the technology for creating and distributing that data. Even the private sector has not yet implemented cost-effective commercial solutions because of the complexity and resource requirements of such a task.

Twitter chief executive Dick Costolo this year announced that the company is working on providing Twitter users with access to all their own tweets. In a July 24 *New York Times* interview he also addressed the question of a search engine enabling access to all tweets, saying, "It's two different search problems. It's a different way of architecting search, going through all tweets of all time. You can't just put three engineers on it."

In the near term, the Library is working to develop a basic level of access that can be implemented while archival access technologies catch up. The Library will consult with congressional researchers and scholars to inform this process. These efforts are ongoing and a priority for the Library. Potential scenarios include public-private partnerships and leveraging private sector investment and capacity.

Recently, senior Library officials met with Gnip senior management in Washington to explore the possibility of developing a research- and scholarship-focused interface to the archive using Gnip's existing historical Twitter product offerings.

The Library continues on a daily basis to build and preserve this important archive, with the expectation that it will be accessible to researchers on premises.

The Library is managing this collection in keeping with part of its mission to acquire, preserve and provide access to a universal collection of knowledge and the record of America's creativity for Congress and the American people. The Library looks forward to continued collaboration with the private sector and the research community as we continue to maintain and build the collection and work toward making this resource accessible for scholarship in a comprehensive, useful way.

APPENDIX 4

List of Interviews conducted

Background interviews with Standards Experts

During investigation of the viability and utility of conduction this study, the researcher conducted a series of unstructured interviews with experts and participants in the arena of media and television in Croatia.

Name	**Affiliation**	**Function**	**Date of Interview**
Ana Habajec	Fremantle Media Hrvatska	Managing Director	April 2014
Mario Kojundžić	RTL Televizija	Digital Project Manager	April 2014
Ivan Lovreček	RTL Hrvatska	Deputy CEO	April 2014
Alida Žorž Miketek	AGB Nielsen Croatia	General Manager	November 2014
Dražen Oreščanin	Poslovna Inteligencija	General Manager	March 2014

CURRICULUM VITAE

Beatrice Züll is a media scientist with extensive experience in the television industry and media sciences. She was born in 1969 in Boppard (a. Rhein), Germany, and grew up in Stuttgart. She went to secondary school in Stuttgart from 1979 to 1988. Studying business and economics (Master's degree) at the University of Siegen from 1989 to 1995, she wrote her dipoma thesis with the title "Possibilities of film financing with special consideration of risk distribution and risk coverage". After her Master's studies, she worked for fifteen years in leading positions for the international television industry. From 1996 to 1999 she worked for companies within the Kirch Group (Taurus Film and ProSieben Media AG) in film financing and television program licensing. In 1999 she transferred to the RTL Group where she was vice president for ten years; and later became senior vice president and was in charge of international television program acquisitions. She represented RTL Television on all national and international program markets and at film festivals such as Cannes Film Festival, MIP TV, Deauville American Film Festival, May Screenings, New York Upfronts, Berlinale, and BBC Showcase. Heading a team of ten employees, she was in charge of negotiating TV program deals with all major studios in the US (The Walt Disney Company, Warner Bros. Entertainment, Paramount Pictures, Sony Pictures, 20th Century Fox, Dreamworks) and other independent film and television production companies. She was nominated as a jury member of the international TV Emmy Awards in the year 2000.

In 2010 Beatrice Züll moved to Croatia and in 2011 joined the PhD Program at the University of Zagreb, at the faculty of Information and Communication Sciences. She became a lecturer at the Academy of Drama and Arts (ADU) in Zagreb in 2012 where she created two new curriculums – television and social media management – for both bachelor and Master's levels. She has published in leading journals in the field of communication and internet research, such as the International Journal of Computers and Communications and WSEAS Press (World Scientific and Engineering Academy and Society). In 2012 and 2013 she joined the Comparing Media Systems meetings at the Inter University Centre in Dubrovnik. In 2012 she presented her research topic "Impact of Social Media on Television" at the 8th Dubrovnik Media Days. At the WSEAS European Computing Conference in 2013 she was chair of the session "Social Implications of Modern Technologies" and presented her research topics "Social Media and Sentiment Analysis", "Social Media and Television", "Internet Freedom" and "Protecting the Right to be Forgotten".

In 2013 she joined the European Cooperation in Science and Technology (COST) (Transforming Audiences, Transforming Societies: working group 2) and she was a member of ECREA (European Communication Research and Education Association). Her research is focused on transforming media industries, transforming audiences; as well as the impact of big data and social media analysis on media management processes. Given her long term working and research experience Beatrice Züll has an extensive network in the media industry and media sciences.

ŽIVOTOPIS

Beatrice Züll je medijski stručnjak s dugogodišnjim iskustvom u televizijskoj industriji i medijskim znanostima. Rođena je 1969. godine u Boppardu (na Rajni) u Njemačkoj, te je odrasla u Stuttgartu.

Srednju školu pohađala je od 1979. do 1988. godine u Stuttgartu. Na Sveučilištu u Stuttgartu studirala je poslovnu ekonomiju od 1989. do 1995. godine, te je stekla titulu magistra. Diplomirala je s dipliomskim radom pod nazivom "Möglichkeiten der Finanzierung von Filmprojekten unter besonderer Berücksichtigung der Risikoverteilung und –absicherung". Nakon završetka diplomskog studija petnaest je godina radila na visokim pozicijama u internacionalnoj televizijskoj industriji. Od 1996. do 1999. godine bila je zaposlena u kompanijama Kirch Grupe (Taurus Film i ProSieben Media AG), gdje je radila u domenama financiranja filmova i licenciranja televizijskog programa. Godine 1999. prešla je u RTL Grupu, gdje je deset godina obnašala dužnost potpredsjednice. S vremenom je postala viša potpredsjednica, te je bila zadužena za internacionalnu akviziciju televizijskog programa. Predstavljala je RTL Televiziju na svim nacionalnim i internacionalnim tržištima programa, te na filmskim festivalima koji su uključivali Cannes Film Festival, MIP TV, Deauville American Film Festival, May Screenings, New York Upfronts, Berlinale i BBC Showcase. Vodila je tim od deset zaposlenika, te je vodila pregovore oko TV programa sa svim glavnim američkim studijima (The Walt Disney Company, Warner Bros. Entertainment, Paramount Pictures, Sony Pictures, 20th Century Fox, Dreamworks) i različitim nezavisnim televizijskim i filmskim produkcijskim kućama. 2000. godine nominirana je za člana žirija internacionalne televizijske nagrade Emmy.

Godine 2010. Beatrice Züll preselila je u Hrvatsku, te je 2011. godine upisala doktorski studij na Fakultetu za informacije i komunikaciju Sveučilišta u Zagrebu. Godine 2012. počela je predavati na Akademiji dramskih umjetnosti (ADU) u Zagrebu, gdje je pokrenula programe menadžmenta u televiziji i društvenim medijima za preddiplomski i diplomski studij.

Objavljivala je u vodećim časopisima za komunikaciju i internetska istraživanja, među kojima su International Journal of Computers and Communication i WSEAS Press (World Scientific and Engineering Academy and Society). Godine 2012. i 2013. sudjelovala je na sastancima Comparing Media Systems-a koja su se održavala na Inter University Centre-u u Dubrovniku. Godine 2012. predstavila je temu svog istraživanja pod nazivom "Impact of Social

Media on Television“ na 8. Dubrovačkim Danima Medija. Godine 2013. vodila je program pod nazivom “Social Implications of Modern Technologies“ u sklopu konferencije WSEAS European Computing Conference, te je predstavila svoja istraživanja pod imenom „Social Media and Sentiment Analysis“, “Social Media and Television“, “Internet Freedom“ i “Protecting the Right to be Forgotten“.

Godine 2013. pridružila se COST-u (European Cooperation in Science and Technology), radna grupa 2: Transforming Audiences, Transforming Societies. Također, bila je članica ECREA-e (European Communication Research and Education Association). Njezina istraživanja fokusirana su na transformaciju medijskih industrija i transformaciju publike, kao i na utjecaj analiza podataka i društvenih medija na procese medijskog menadžmenta. Zbog svog dugogodišnjeg radnog i istraživačkog iskustva, Beatrice Züll povezana je sa širokom mrežom medijske industrije i medijskih znanosti.