Samuel Neaman Institute
for National Policy Research

Society

Education

Economy

Science & Technology

Environment & Energy

Long-term Planning

Industry & Innovation

Physical Infrastructure

Health

Human Capital

Higher Education

An Innovative Approach for Measuring the Digital Divide in Israel: Digital Trace Data as Means for Formulating Policy Guidelines

Professor Sheizaf Rafaeli
Dr. Eran Leck
Dr. Yael Albo
Yael Oppenheim
Dr. Daphne Getz

July
2018

Samuel Neaman Institute
for National Policy Research

An Innovative Approach for Measuring the Digital Divide in Israel: Digital Trace Data as Means for Formulating Policy Guidelines

Final Report

The Samuel Neaman Institute for National Policy Research

Researchers:
Professor Sheizaf Rafaeli
Dr. Eran Leck
Dr. Yael Albo
Yael Oppenheim
Dr. Daphne Getz

Haifa, July, 2018

ABOUT THE SAMUEL NEAMAN INSTITUTE

The Samuel Neaman Institute was established in 1978 in the Technion at Mr. Samuel Neaman's initiative. It is an independent multi-disciplinary national policy research institute. The activity of the institute is focused on issues in science and technology, education, economy and industry, physical infrastructure and social development which determine Israel's national resilience.

National policy research and surveys are executed at the Samuel Neaman Institute and their conclusions and recommendations serve the decision makers at various levels. The policy research is conducted by the faculty and staff of the Technion and scientists from other institutions in Israel and abroad and specialist from the industry.

The research team is chosen according to their professional qualifications and life achievements. In many cases the research is conducted by cooperation with governmental offices and in some cases at the initiative of the Samuel Neaman institute and without direct participation of governmental offices.

So far, the Samuel Neaman Institute has performed hundreds of exploratory national policy research projects and surveys that serve decision makers and professionals in economy and government. In particular the institute plays an important leading role in outlining Israel's national policies in science, technology and higher education.

Furthermore, the Institute supports national projects, such as the Ministry of Industry, Trade & Labor clusters - the MAGNET program in nano-technologies, media, optics and communication, chemistry, energy, environmental and social projects of national importance. The institute organizes also comprehensive seminars in its leading fields of research.

The Samuel Neaman Institute's various projects and activities can be viewed at the Institute website.

The chairman of Samuel Neaman Institute is Professor **Zehev Tadmor** and the director is Professor **Moshe Sidi**. The institute operates within the framework of a budget funded by Mr. Samuel Neaman in order to incorporate Israel's scientific technological economic and social advancement.

The opinions and conclusions expressed in this publication are those of the authors and do not necessarily reflect the opinion of the Samuel Neaman Institute.

Table of Contents

List of Figures

List of Tables

List of Annexes

Acknowledgments

This research was supported by the **Ministry of Science and Technology (MOST).** We thank MOST for their generous financial support which made this research possible.

We gratefully acknowledge the productive cooperation and assistance of Mr. Nati Yaakobi, CEO of Ifat Media Advertising Monitoring, and the Ifat team, who provided us with comprehensive and insightful data. The specially tailored dataset developed by the Ifat Media Monitoring team proved to be priceless for the research and vastly contributed to its enrichment.

We would like to express our deep gratitude to Dr. Einat Orr, CTO at SimilarWeb, who opened for us the doors of SimilarWeb's fascinating world of tools and data. We would also like to thank Mr. Felix Vaisman, Senior Team Lead at SimilarWeb, for his kind and patient help with problem solving.

We would like to thank the Buzzilla team for their technical assistance with web-conversation extraction from the Buzzilla tool.

We thank Dr. Nitza Barkan, Senior Consultant and a lecturer at the Statistics Department at the University of Haifa, and Prof. Ron Kenett, Senior Research Fellow at the Samuel Neaman Institute for their statistical consulting and methodological insights. We also thank Dr. Eitan Adres, Senior Research Fellow at the Samuel Neaman Institute, for sharing his expertise on small and medium businesses.

We would like to express our sincere gratitude to Dr. Joel Lanir, a Senior Lecturer at the Information Systems Department and Head of the HCI and Visualization Lab at the University of Haifa for his valuable comments on the data visualization chapter.

We thank our colleague Ms. Frida Elek, a social worker and a PhD student at the University of Haifa, for her precious help regarding social rights realization. We also thank Lottem Shmuel for help in the manual data categorization process.

We would also like to thank Ms. Lydia Jiryes and Ms. Samira Mahameed for their kind help in directing us to relevant social media channels in the Arab language.

We thank Mr. Golan Tamir, Information Systems Manager at the Samuel Neaman Institute for his technical support.

Finally, we would like to express our deep gratitude to Dr. Adi Libsker-Hazut - Research, Planning and Academic Projects Coordinator at the Center for Internet Research in the University of Haifa, for her precious administrative assistance.

Hebrew Executive Summary

תקציר מנהלים

במהלך שני העשורים האחרונים, נצפו שינויים משמעותיים ומהירים בהתפשטותן של טכנולוגיות התקשורת והמידע. השימוש ההולך וגובר באינטרנט משפיע באופן ניכר על מרבית תחומי החיים, תוך כדי שהוא הולך ומשנה את הדרך בה בני האדם מתקשרים, צורכים מידע ומנהלים את פעילויות היומיום שלהם. עם זאת, מתברר כי טכנולוגיות המידע לא אומצו במידה שווה על ידי כל חלקי האוכלוסייה, וכי קיימים הבדלים בגישה לאינטרנט, באופני השימוש בו ובסוג התוכן המקוון הנצרך על ידי קבוצות סוציו-דמוגרפיות שונות. המחקר של תופעה זו, המכונה "הפער הדיגיטלי", הופך חשוב יותר ויותר בשנים האחרונות לצרכי התווית והערכת מדיניות לצמצום פערים דיגיטליים.

שיטות וכלים נפוצים להערכת הפער הדיגיטלי כוללים סקרים, ראיונות מובנים ושאלונים. שיטות "דיווח עצמי" אלו, בעוד שהן חשובות ומועילות מאוד, סובלות ממספר חולשות. בין חסרונותיהן ניתן למנות את היותן פולשניות, יקרות, בלתי ניתנות לשכפול או ניתנות לשחזור, אינן מאפשרות לעיתים קרובות לערוך ניתוח על בסיס גיאוגרפי ברמת הבחנה גבוהה וחשופות להטיות דגימה משמעותיות.

בשנים האחרונות הולך וגובר השימוש בכלי מחקר הנסמכים על נתונים המתעדים את ההתנהגות האנושית המקוונת (למשל לוגים של שימוש באינטרנט, שיחות מקוונות ועוד). היתרון העיקרי בשימוש בנתונים אלו, המכונים "עקבות דיגיטליים", נובע מהיותם נתוני אמת ולא נתוני דיווח עצמי. בפרויקט זה מוצגת גישה חדשנית ובלתי פולשנית לזיהוי, איסוף, ניתוח והדמיה של הפער הדיגיטלי, תוך שימוש בנתוני עקבות דיגיטליים. המטרה העיקרית של המחקר היא לספק את הבסיס התיאורטי והמעשי למדידה והערכה של הפער הדיגיטלי באמצעות נתונים אלו.

במסגרת המחקר נעשה שימוש בשישה מקורות וכלים שונים של נתוני עקבות דיגיטליים, המתבססים על מאפיינים סוציו-דמוגרפיים ומרחביים שונים. מקורות אלה כוללים נתוני לוגים של פאנל גולשים, מערכים של נתוני גלישה אגרגטיביים, נתונים שהופקו באמצעות שימוש בתוכנות הסוקרות שיחות מקוונות, כלי ניטור של שימוש באתרים ספציפיים וכן כלי ניטור של מונחי חיפוש במרחב האינטרנטי. מערכי הנתונים הגולמיים נוקו, עובדו, קודדו ונותחו הן בנפרד והן במשולב.

במסגרת המחקר, יושמה גישת טריאנגולציה אשר כללה שילוב ויישום של מספר שיטות וכלים, במטרה לשפר את יכולת ניתוח תופעת הפער הדיגיטלי באמצעות הצלבת ממצאים. השימוש במתודולוגיה הודגם באמצעות חקר מקרה (case study) של בחינת פערים דיגיטליים בתחום של מימוש זכויות חברתיות. ממצאי חקר-המקרה הוצגו באמצעות סיפורי נתונים הנתמכים בהדמיה חזותית של הפערים שתוארו. המחקר דן במרחב הפתרונות הוויזואליים של נתוני עקבות דיגיטליים בהקשר של הערכת פער דיגיטלי, תוך הדגשת מאפייני הנתונים וביניהם ריבוי ממדים, תלות בממד הזמן ולעיתים אף ריבוי מקורות שאינם אינטגרטיביים.

המחקר עשה שימוש במגוון רחב של שיטות סטטיסטיות תיאוריות וכמותיות וכן בכלי מחקר איכותניים, הכוללים ניתוח טקסטואלי של דיונים מקוונים. הממצאים סיפקו הוכחת היתכנות ליעילות השימוש בנתוני עקבות דיגיטליים למטרת חקר הפער הדיגיטלי, וליכולתם של כלים בלתי פולשניים להחליף שיטות דיווח עצמי במשימה זו. ממצאי המחקר מצביעים על קיומם של פערים דיגיטליים, כפי שהם משתקפים על ידי **נפח השימוש** (המספר הממוצע של הביקורים באתרי אינטרנט והתפלגותם), **מגוון השימוש** (מספר קטגוריות האתרים בהן ביקר המשתמש) ו**תכני השימוש** (סוג הפעילות המקוונת או התוכן הנצרך) אשר נמצאו כמדד המשמעותי ביותר מבין השלושה בשיקוף פערים דיגיטליים.

מבחינת ***נפח השימוש***, נמצא כי הוא גבוה יותר בקרב גברים בהשוואה לנשים. כמו כן, נמצאו הבדלים משמעותיים בהיקף השימוש בין דוברי עברית, ערבית ורוסית. ממוצע הביקורים בקרב דוברי העברית היה גדול פי שניים מדוברי ערבית וגבוה פי 2.4 מאשר דוברי רוסית. מהממצאים עולה כי קיימים הבדלים משמעותיים בנפח השימוש בין משתמשים מאזור המרכז (מחוז תל אביב) לבין הפריפריה (מחוזות צפון ודרום), כאשר נפח השימוש (מספר הביקורים הממוצע) המאפיין את תושבי המרכז גדול פי חמישה מזה של המשתמשים בפריפריה.

מבחינת ***מגוון התוכן***, התברר כי צריכת התוכן של גברים היא מגוונת יותר מזו של נשים. כמו כן, נמצאו הבדלים משמעותיים בין מרחבים גיאוגרפיים שונים, כך שמשתמשים במטרופולין תל אביב ובמטרופולין ירושלים מאופיינים ברמת מגוון הגבוהה באופן מובהק בהשוואה למשתמשים מאזורים אחרים. נמצא כי מגוון התוכן עולה בהתאם לרמת ההשכלה, כאשר משתמשים בעלי השכלה על-תיכונית ומעלה הראו רמת גיוון גבוהה משמעותית מזו של בעלי השכלה תיכונית או נמוכה יותר. באופן מפתיע, משתמשים חרדים הציגו את הרמה הגבוהה ביותר של גיוון והם נבדלים סטטיסטית משאר הקבוצות.

מבחינת ***שימוש בתוכן***, ממצאי המחקר חושפים רמה גבוהה של תאימות בין תוצאות מחקר זה הנסמך על עקבות דיגיטליים לבין ממצאים שונים המדווחים בספרות המתבססים על מקורות דיווח עצמי:

- פערים דיגיטליים מגדריים נמצאו משמעותיים בקטגוריות התוכן הבאות: מידע וחיפוש; בידור; פיננסים; היכרויות (תחומים הנשלטים על ידי גברים) ובריאות (תחום נשלט על ידי נשים).
- הבדלים בין דוריים משמעותיים בהתנהגות המקוונת זוהו בקטגוריות התוכן הבאות: דואר אלקטרוני; בריאות; קניות מקוונות (תחומים הנשלטים על ידי קבוצות גיל מבוגר) ובידור (תחום הנשלט על ידי קבוצות גיל צעירות).
- פערים דיגיטליים מבוססי רמות השכלה זוהו בקטגוריות השימוש בתכנים הבאים: מימוש זכויות; חדשות; עבודה, קריירה, מחקר וחינוך; פיננסים (תחומים הנשלטים על ידי משתמשים בעלי רמות השכלה גבוהות יותר) ובידור (מוסיקה, וידאו ומשחקים וכו '); כלי תקשורת, מסרים מיידיים, צ'אטים, רשתות חברתיות, והימורים (תחומים הנשלטים על ידי משתמשים בעלי רמות השכלה נמוכות יותר).
- הבדלים משמעותיים בהתנהגות המקוונת בין משתמשים בעלי הכנסה גבוהה לבין משתמשים בעלי הכנסה נמוכה זוהו בקטגוריות התוכן הבאות: בידור (מוסיקה, וידאו ומשחקים וכו '); כלי תקשורת,

מסרים מיידיים, צ'אטים ורשתות חברתיות (תחומים הנשלטים על ידי משתמשים עם רמות הכנסה נמוכות יותר) ותיירות מקוונת (תחום הנשלט על ידי משתמשים בעלי רמות הכנסה גבוהות יותר).

ממצאי חקר-המקרה של מימוש הזכויות המחישו את הכדאיות של הערכת הפער הדיגיטלי באמצעים של טריאנגולציית נתונים דיגיטליים. להלן התובנות שעלו במסגרת חקר-המקרה:

- נשים נוטות להיות מעט פחות פעילות באינטרנט בהשוואה לגברים ביחס למימוש זכויות. העניין במימוש הזכויות הולך ופוחת עם הגיל, כאשר משתמשים צעירים פעילים משמעותית יותר ממשתמשים מבוגרים.
- ניכר כי חלקם של המשתמשים הזקוקים לגורם מתווך לצורך מימוש זכויותיהם גבוה יחסית בקרב האוכלוסייה המבוגרת.
- בחינת נתוני עקבות דיגיטליים בנושא זכויות מעניקה יכולת הבנה של הליך שיום הזכויות בקרב הציבור, המהווה שלב הכרחי לצורך מימושן. ממחקר קודם עולה כי יכולתו של משתמש לספק שם מדויק לזכות ספציפית הינה שלב חיוני לצורך מימוש אותה הזכות. בחקר המקרה הודגמה אפשרות הניתוח של המונחים הלשוניים השגורים בהקשר של הזכויות החברתיות, ואולי אף חשוב מכך בחינת המונחים בהם **לא** נעשה שימוש.
- פייסבוק הינו ערוץ המדיה החברתית הפופולרי ביותר בנושא מימוש זכויות, בעוד טוויטר וזירת הבלוגים הם הכי פחות פופולריים למטרה זו.
- פרסומים בתקשורת מרכזית (ערוצי חדשות) מייצרים עניין ציבורי ומעורבות ותורמים לשיח הציבורי בתחום של מימוש זכויות.

ממצאיו ותוצאותיו של מחקר זה יכולים לספק למשרדי הממשלה בישראל ולקהילה המחקרית תובנות מרכזיות הדרושות לשם גיבוש מדיניות ציבורית בתחום הפער הדיגיטלי, כמו גם לקחים מתודולוגיים ופרוצדוראליים אשר יכולים לשמש למחקר מתקדם בתחום העקבות הדיגיטליים**.**

ההמלצות לגורמי ממשל הן כדלקמן:

אנו ממליצים למשרד המדע והטכנולוגיה לפעול ליצירת פרוטוקול אשר יסדיר ויגדיר את השימוש בנתוני עקבות דיגיטליים. על הפרוטוקול להגדיר הנחיות ברורות עבור: איסוף, ניטור וכריית נתונים ממקורות מקוונים; אנונימיזציה של מידע אישי מטעם בעל הנתונים; נהלים לעיבוד נתונים, איחוד וקישור של נתוני עקבות דיגיטליים ממקורות מרובים; הנחיות לגבי הצגת הנתונים (מטעם החוקר); בנייה ותחזוקה של מאגרי עקבות דיגיטליים (עם או באמצעות גופים כגון הספרייה הלאומית או ארכיון המדינה); שימוש של צד שלישי; הקנסות שיוטלו על החוקר במקרה של הפרת תנאי החוזה.

אנו ממליצים למשרדי הממשלה, לספקי השירותים הממשלתיים המקוונים (בנקים, ספקי שירותי בריאות, אוניברסיטאות וכיו"ב) ולגופים בחברה האזרחית (עמותות להנגשת מידע ציבורי):

- להעלות מודעות, בעיקר בקרב נשים, צעירים, מעוטי הכנסה ובעלי השכלה נמוכה, **לחשיבות של רכישת ידע בתחום הפיננסים ובתחום הדיור באמצעים מקוונים.**
- להעלות מודעות, בעיקר בקרב גברים, צעירים, מעוטי הכנסה והאוכלוסייה החרדית באשר ליתרונות ב**ביצוע פעולות מקוונות בתחום הבריאות**.
- להעלות מודעות, בעיקר בקרב מבוגרים, מעוטי הכנסה ובעלי השכלה נמוכה באשר ליתרונות בשימוש **באתרי ממשלה (e-gov) ואתרי הרשויות המקומיות המספקים שירותים מקוונים לאזרח.**
- לבצע מעקב שיטתי אחר התנהגות הגולשים באתרי הזכויות והממשל המקוון על מנת לזהות זכויות "חמות" המושכות אליהן קהל רב וזכויות אשר המידע אודותיהן אינו מגיע לקהל הזכאים, וכן על מנת לזהות מגמות ושינויים עונתיים או תקופתיים בהתנהגות המשתמשים באתרים אלו.
- להעלות מודעות, בעיקר בקרב נשים, בני נוער, חרדים ומעוטי הכנסה באשר לחשיבות של חיפוש מידע וביצוע פעולות מקוונות בהקשר של **זכאות להטבות ומימוש זכויות.**
- המחקר חשף את חשיבותו של שימוש מדויק במונחי חיפוש לצורך שליפת מידע בנושא מימוש זכויות חברתיות. אנו ממליצים לגוף הממשלתי הרלוונטי (המוסד לביטוח לאומי) להעמיק את בחינת השימוש שעושות (או לא עושות) אוכלוסיות שונות במידע הקיים אודות זכויות, על מנת לשפר את התאמתם והנגשתם של האתרים הרלוונטיים (לדוגמה www.btl.gov.il).
- עידוד השימוש ברשתות חברתיות ובפורומים (בעיקר בקרב האוכלוסייה החרדית) ע"י משרדי ממשלה לצורך הפצת מידע ויצירת מודעות ציבורית בתחום של ***מימוש זכויות***, במטרה להגיע ישירות לאוכלוסיות מוחלשות.
- לבצע ניטור וניתוח שוטפים של השיח בנושא מימוש זכויות בעמודי הפייסבוק של משרדי הממשלה וספקי השירותים הממשלתיים הרלוונטיים, בתקשורת המרכזית ובפורומים השונים. זאת על מנת להעמיק את הבנת הצרכים של הקהלים הפעילים בשיח כמו גם על מנת לזהות את הקהלים שאינם פעילים בשיח.
- העלאת המודעות לחשיבות וליתרונות הקיימים בפעילויות של ***חינוך מקוון ולמידה מקוונת***, בעיקר בקרב מבוגרים ומעוטי הכנסה.
- העלאת המודעות וטיפול בבעיית ***פעילות הימורים מקוונת***, בעיקר בקרב גברים, צעירים, מעוטי הכנסה ובעלי השכלה נמוכה.

המלצותינו לקהילת החוקרים המשתמשים בנתוני עקבות דיגיטליים הן כדלקמן:

- קידום ופיתוח של מתודולוגיות לטריאנגולציה של נתונים וכלים שיתרמו לשיפור מהימנות הנתונים ולהבנת התופעה הנחקרת (לדוגמה פער דיגיטלי).
- קידום ופיתוח מתודולוגיות מחקר לניתוח נושאי של תוכן מקוון במאגרי עקבות דיגיטליים גדולים באמצעות טכנולוגיות בינה מלאכותית ולמידת מכונה. מאמץ זה קשור לעיבוד שפה טבעי (Natural Language Processing), תחום מאתגר במיוחד עבור השפה העברית.

- פיתוח ושיפור מתודולוגיות מחקר לאיחוד סקרים אינטרנטיים עם נתוני עקבות דיגיטליים (שיפור ייצוגיות הדגימה, התוכן וכיו"ב), על מנת להעמיק את ההבנה של התנהגות מקוונת גלויה של משתמשים.
- הידוק הקשרים עם עמותות להנגשת ידע ציבורי, עמותות לקידום מימוש זכויות וכן עם חברות מסחריות העוסקות בניטור וניתוח של נתוני עקבות דיגיטליים לצורך שיתופי פעולה מחקריים.

Executive Summary

Over the past two decades, vast and rapid changes have been witnessed in the use and diffusion of information technologies. The introduction and growing use of the internet has exerted a substantial impact on everyday life, changing the way humans interact, consume information and conduct their daily activities. However, the adoption of information technologies has not been equally met by all members of society, resulting in gaps in access, usage and the type of on-line content consumed across socio-demographic, economic and spatial landscapes. The study of this phenomenon, known as the **digital divide**, is becoming increasingly important in recent years for policy purposes.

Commonly used methods and tools for the evaluation of the digital divide include surveys, structured interviews, open questionnaires and indicator analysis. These “self-report” methods, while very important and useful, are prone to several weaknesses. They are obtrusive, costly, unreplicable, have very little granularity with respect to regional analyses and are subjects to real sampling bias.

In this project, an innovative and novel approach for identifying, collecting, analyzing and visualizing the digital divide is presented, using unobtrusive methods. The main goal of the research is to supply the theoretical and practical underpinning for measuring and evaluating the digital divide using digital trace data.

In the framework of the study, six different digital trace data sources, parsed with reference to socio-demographic and spatial attributes, were used to analyze online user behavior, with the specific aim of studying digital gaps. The raw datasets were cleaned, processed, coded and analyzed, both on an individual and on a triangulation basis. The triangulation approach involved the combination and application of several methods and tools with the specific aim of facilitating the understanding of the digital divide phenomenon. This methodology was demonstrated by a case-study that investigated and analyzed digital gaps in the rights realization domain and involved the use of data stories that supplied systematic guidance for researching and understanding these divides. The data-driven stories were subsequently portrayed by data visualization. The design space of data visualization of trace data in the digital divide context was discussed, highlighting its multi-dimensional, time-oriented and multi-source characteristics. The research findings were presented using a wide range of descriptive and quantitative statistical methods as well as qualitative tools, involving textual analysis of on-line discussions.

The results of the research provide both a proof of concept and important insights regarding the use of digital trace data in the study of the digital divide and as to the ability of unobtrusive tools to replace self-report methods in this task. The findings of the research pointed out the existence of digital gaps, as reflected by ***usage volume*** (number of visits/distribution of visits), ***variety*** (the number of different website categories visited by the user) and ***content usage*** (type of on-line activities), with the latter category being the most significant in terms of gaps out of the three.

In terms of usage volume, male users were found to exhibit higher usage volume than female users. Significant differences in usage volume were also observed between Hebrew, Arabic and Russian speakers. The usage volume among Hebrew speakers was two times larger than Arabic speakers and 2.4 times larger than native Russian speakers. Stark spatial differences in usage volume were found between users from the Core ("Center") region (Tel Aviv District) and the country's periphery (North and South Districts), with the usage volume characterizing Core residents being five times larger than the one characterizing users from the Periphery.

In terms of Internet content diversity, male users were found to be more diverse than female users with respect to internet content consumption. Spatial differences with respect to the diversity level were also found to be significant, with users from the Tel Aviv metropolitan region and the Jerusalem metropolitan region exhibiting the highest diversity levels and statistically differ from users from other regions. The level of diversity was found to rise with education level, where individuals with post-secondary education or higher level education having a substantially higher diversity level than individuals holding secondary or lower level education. Surprisingly, ultra-orthodox users exhibited the highest level of diversity and statistically differ from all other groups.

In terms of content usage, the findings of the research reveal a high degree of compatibility between the results of this digital trace exercise and various findings reported in the literature from self-report sources:

- Digital gaps in online behavior between female and male users was found to be substantial in the following content usage categories: Information and Search; Entertainment; Finance; Dating (dominated by males) and Health (dominated by female).

- Substantial generational differences in online behavior were identified in the following content usage categories: E-mail; Health; On-line shopping (dominated by older age cohorts) and Entertainment (dominated by younger age cohorts).
- Substantial education-based differences in online behavior were identified in the following content usage categories: Government and rights realization; News; Work, career, research and education; Finance (dominated by users with higher levels of education) and Entertainment (music, video and gaming etc.); Communication tools, Instant messaging, chat and social networks and Gambling (dominated by users with lower levels of education).
- Substantial differences in online behavior between high-income users and low-income users were identified in the following content usage categories: Entertainment (music, video and gaming etc.); Communication tools, Instant messaging, chat and social networks (dominated by users with lower levels of income) and Travel and tourism (dominated by users with higher levels of income).

The findings of the rights realization case study demonstrated the feasibility of evaluating the digital divide by means of digital trace data triangulation. The following insights were derived from the various data stories presented in the rights realization case study:

- Women tend to be slightly less active than men with respect to the realization of rights. Rights realization decreases with age, where young users are the most active in the realization of rights and the activity of older users in this respect is substantially lower.
- The share of users requiring mediation for the realization of their rights is substantially higher among older populations.
- Examining digital trace data concerning entitlements facilitates an understanding of the naming procedure (e.g. the ability of a user to provide an accurate name for a specific right, which is required for its realization).
- Facebook constitutes the most popular social media channel for rights realization, while Twitter and Blogs are the least popular.
- News media coverage prompt public interest, active involvement and contribute to public discourse in the rights realization domain.

The findings of this research can supply various government actors in Israel key insights for the formulation of public policy in the digital divide domain. The research outputs could be also of great use for the research community at large, as

they provide valuable methodological and procedural lessons that can be utilized for advanced research in the field of digital traces.

The recommendations are as follows:

We recommend that the Ministry of Science and Technology be active in the formulation of protocols to define and regulate the use of digital trace data.

Such a protocol should set clear guidelines for: Data collection and data mining from on-line sources; The anonymization of personal information on behalf of the data owner; Accepted practice and procedures for data processing, cross referencing and consolidation of digital trace data and survey data from multiple sources; Guidance regarding the presentation of the data (on behalf of the researcher); The construction and maintenance of digital trace repositories (with or through entities such as the National Library or the Israel State Archives-ISA); Third party use; and the penalties that might be imposed on the researcher in case of breaching the contract terms.

We recommend that the relevant government offices, service and data providers of on-line platforms (e.g. banks, e-health and municipal service providers, universities, etc.) and social society actors (e.g. NGOs involved in making online information accessible to the public):

- Raise awareness and enhance education, especially among women, young adults, lower income and lower education populations as to the ***importance of on-line financial education and knowledge of the housing market***.
- Raise awareness and enhance education, especially among men, young adults, lower income and ultra-orthodox populations as to the ***benefits of using and conducting e-health activities***.
- Raise awareness and enhance education, especially among older adults, lower income and lower education populations as to the benefits of using ***e-gov and on-line municipal services***.
- Raise awareness and enhance education, especially among women, young adults, ultra-orthodox and lower education populations as to the importance of searching information and conducting on-line transactions with regards to ***entitlement benefits and rights realization.***
- The research has exposed the importance of defining and using accurate search terms in retrieving information (see the naming story). We recommend the relevant

government actor (National Insurance Institute) to learn about the variant use of each right by its users, with the specific aim of better customizing relevant websites (e.g. www.btl.gov.il).

- Encourage the use of social networks and blogs (especially among ultra-orthodox population) among government offices in disseminating knowledge and raising public awareness in the domain of ***rights realization***, with the specific aim of targeting deprived populations.
- Raise awareness and enhance education, especially among older adults and lower income population as to the importance and benefits ***of e-education and e-learning activities***.
- Raise awareness and address the problem of increased on-line ***gambling activity*** especially among men, young adults, lower income and lower education populations.

Our recommendations to the *research community are*:

- Promote and develop data triangulation methodologies and tools for the purpose of enhancing data reliability and understanding of the investigated phenomena (e.g. digital divide).
- Promote and develop research methodologies for categorizing internet content using machine learning and AI techniques for large corpus digital trace data. This effort is related to Natural Language Processing (NLP) which is especially challenging with regards to the Hebrew language.
- Develop and improve existing methodologies for consolidating internet panels with digital traces (e.g. representativity of the sample of on-line users, representativity of content, etc.) for the purpose of deepening understanding of overt online user behavior.

Introduction

The digital revolution impacts our daily life, influencing individuals, households and workplaces alike. The vast diffusion of ICT products and services has produced much positive economic and societal spillovers, but has also lead to the creation of demographic (age, gender, ethnic background etc.) and spatial disparities (e.g. core versus periphery) between various population groups. This disparity, known as the digital gap or the digital divide, is a social issue relating to the inequality of access to digital technologies. The OECD (2001) defines the digital divide as “differences between individuals, households, companies, or regions related to the access to and usage of ICT”. The divide may appear due to historical, socioeconomic, geographic, educational, behavioral, or generation factors, or due to the physical incapability of individuals (Cullen, 2001). The literature shows that access to and acquisition of digital technologies and proficiency provides many advantages: from enhanced employment and education opportunities to efficient utilization of public services (McLaren and Zappala, 2002; Rice and Katz, 2003; Losh, 2004; Moon et al., 2010; Shirazi et al., 2010).

The study of digital divide is becoming increasingly important for policy formulation and policy evaluation purposes. The empirical literature relating to the digital divide is abundant. A wide array of qualitative and quantitative methods has been used over the years to measure the scope of the digital divide between countries and various populations groups and to identify the main factors affecting it. These methods include surveys, structured interviews, open questionnaires and indicator analysis. While very useful and important, these “self-report” methods are obtrusive and very costly, thus un-replicable. The traditional methods are prone to several weaknesses, they are, by-and-large one-shot studies, provide little continuous or even benchmark measures, costly, have very little granularity with respect to regional (or other sub-group) analyses and are subject to real sampling and self-report biases.

In this project, we offer an innovative and novel approach for identifying, collecting, analyzing and visualizing unobtrusive digital traces data. We develop and employ this approach for policy formulation and policy evaluation purposes. In the framework of the research, six different digital trace data sources, parsed with reference to socio-demographic and locational attributes, are used to analyze online user behavior, with the specific aim of studying digital gaps.

The research employs a wide range of descriptive and quantitative research methods and tools, including graphs, two-dimensional tables, statistical tests, regression models and a specially tailored normalized index in order to map and analyze digital traces in Israel.

An important contribution of this research is the formulation of a methodology for triangulating various digital trace data sources in order to deepen our understanding of the digital divide phenomenon and to construct more robust methodological tools, allowing evidence-based evaluation of actions. The triangulation methodology is demonstrated by focusing on a rights realization case study in the context of digital divide. Finally, we describe techniques and methodologies relevant to the visualization of digital traces in general and the digital divide in particular.

The report is organized as follows: **Chapter 1** provides the literature overview for this work. It reviews various definitions of the digital divide and discusses common methods and tools (e.g. composite indexes) that are used to evaluate digital gaps at the national and international levels. It also presents an overview of the socio-demographic and content usage attributes of the digital divide. The chapter concludes with a review of the advantages of unobtrusive research methods in the analysis of on-line human behavior and outlines the various digital trace data sources that can be used for this task. ***Chapter 2*** reviews the methodological framework for this work. It presents the research goal and objectives, the motivation and estimated contribution, the research questions, the research population and the research data. A description of the conceptual framework and research work plan concludes this chapter. ***Chapter 3*** describes the research findings. It maps and evaluates the digital divide in Israel using digital trace data. The analysis centers on the evaluation of gaps in three domains: usage volume, variety and content usage. ***Chapter 4*** presents a methodology for triangulating various digital trace data sources. The triangulation methodology is demonstrated by focusing on a rights realization case study in the context of digital divide. ***Chapter 5*** discusses various aspects and insights regarding the visualization of the digital divide and presents an example for digital divide visualization using “off-the-shelf” tools. ***Chapter 6*** presents a summary and critical discussion of the research findings. ***Chapter 7*** concludes the report with a review of the research limitations, its methodological, theoretical and practical contributions and provides policy implications for stakeholders and the research community.

Chapter 1: Literature Review

The term “digital divide”, originally coined in the early 1990’s (Cruz-Jesus, 2012), typically relates to sociodemographic differences in the use of information and communication technology (Vehovar et al., 2006). The term refers to the gap between individuals, households, businesses and geographic areas at different socio-economic levels, with regards to their access to information and communication technologies (ICTs) and to their use of the Internet for a wide variety of activities (OECD, 2001). The mitigation of digital gaps is seen by many countries as a moral and social interest (raising personal welfare, alleviation of social gaps, promotion of equal opportunities among various population groups), as an economic interest (as means for achieving a competitive advantage) and as a political interest (a strategy for promoting and safeguarding national resilience) [Rafaeli et al., 2013].

Since the mid 1990’s, a voluminous body of literature has accumulated on the digital divide. The literature distinguishes between two main approaches or dimensions for analyzing and measuring the digital divide: internal or domestic divide which measure disparities within a country and international or cross-country comparison which evaluates the gaps between countries on an aggregated level. The measurement of domestic digital divides focuses on the level of ICT access and use to highlight the gaps between groups of people, whether these people are grouped by socio-economic status, geographic location or other characteristics. Cross-country comparisons of the digital divide mostly rely on performance evaluations based on comparative rankings and composite indices (Petrović et al., 2012; Rafaeli et al., 2013).

International Assessment of Digital Divide

Comparing country performances to identify evolutionary trends and establish benchmarks is a common practice in a wide range of fields (e.g. environment, economics and technological development). Such comparisons are often performed by introducing composite indexes (CIs), calculated through the aggregation of individual indicators, reproducing quantitative or qualitative measures of factors with the aim of representing the relative position of a country on a conceptual space (OECD 2011).

The development and use of internationally comparable and reliable ICT indicators for measuring e-readiness and the digital divide are important for policy makers and statistical agencies alike. Since the year 2000, numerous indices aimed at measuring the digital divide were developed (Rafaeli et al., 2013). Two of the most important indices, still in use today are the Networked Readiness Index (NRI) and the Digital Access Index (DA).

The World Economic Forum’s NRI index measures, on a scale from 1 (worst) to 7 (best), the performance of 139 economies in leveraging information and communications technologies to boost competitiveness, innovation and well-being. It measures the capacity of countries to leverage ICTs for increased competitiveness and well-being. The NRI is composed of 4 main sub-indices: environment, readiness, usage, and impact sub-index. Under the environment sub-index, the political, regulatory, environment, business and innovation environments are evaluated. Infrastructure, affordability and skills are assessed in the readiness sub-index, while the impact index measures both economic and social impacts of higher ICT usage. As for the usage sub-category, it considers individual, business and government usages of ICT[1].

ITU’s ICT Development Index (IDI), which has been published annually since 2009, is a composite index that combines 11 indicators into one benchmark measure. It is used to monitor and compare developments in information and communication technology (ICT) between countries and over time. The IDI is built around three fundamental vectors that impact a country's ability to access ICTs: access, use and skills. The IDI 2015 has been calculated for 167 economies where European countries were among the highest ranked, with the exception of the Republic of Korea, being on top. The Index is designed to be global and to reflect changes taking place in countries at different levels of ICT development. It therefore relies on a limited set of data which can be established with reasonable confidence in countries at all levels of development[2].

Using CIs to represent a complex phenomenon provides, in general, advantages and limitations. According to the OECD (2011), the main advantages of CIs are their ability to summarize complex multidimensional phenomena with a view to support decision makers, to assess the progress of countries over time, to facilitate communication with the general public, to promote accountability and to enable users to compare complex dimensions effectively. At the same time, the use of CIs, such as the NRI and IDI, for measuring the digital divide have many notable drawbacks. One of the most severe critiques was raised by Van Dijk (2006), who argues that the attempt to measure the digital divide suffers from

[1] https://www.weforum.org/
[2] http://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2015/methodology.aspx

a lack of adequate theoretical framework, as these indices only emphasize structural factors such as differences in income, education, age, gender, and ethnicity, while not addressing the deeper social, cultural and psychological factors responsible for instigating inequalities (Bruno et al. 2010). Fuchs (2009) criticizes the choice of the indicators in the current indices as they reduce the role of socioeconomic, political and cultural factors and focus mainly on factors relating to technological access and use. Barzilai-Nahon (2006), Menou and Taylor (2006) and James and Versteeg (2007) point out that the choice of the aggregation methodology of individual indicators, data standardization and normalization operations are responsible for significant biases (Bruno et al. 2010). CIs may also send misleading policy messages if poorly constructed or misinterpreted, lead to simplistic or inappropriate policy conclusions. In addition, the selection of indicators and their weights are sensitive to the subject of political dispute. As the variables in these indices are mainly based on national surveys, they are also vulnerable to self-report bias (OECD, 2011).

Segev and Ahituv (2010) conducted a cross-country (international) evaluation of the digital divide using digital trace data. The authors developed and employed an innovative methodology to examine and assess the digital divide in information uses, looking at the extent of political searches and their accuracy and variety. Their findings indicate that some countries, particularly Germany, Russia, and Ireland, display greater accuracy of search terms, diversity of information uses, and socio-political concern.

Domestic Assessment of Digital Divide

Most of the domestic studies conducted since the mid 1990's were questionnaire based and highlighted the differences between various population groups, the accessibility level to ICT and the extent of use in ICT products and services (Rafaeli et al., 2013). One of the first scholarly papers on the digital divide phenomenon was carried out by Katz and Aspden (1997) who analyzed the motivations for and the barriers to internet usage using a national random telephone survey in the United States. Their survey showed evidence of a digital divide - Internet users being generally wealthier and more highly educated, and blacks and Hispanics disproportionately unaware of the Internet. Subsequent studies, focusing on the internal digital divides within a country, have identified one or a few variables that influence measures of digital divide such as awareness, access, attitudes or application (Barzilai-Nahon, Rafaeli and Ahituv, 2004; Barzilai-Nahon, 2006). A partial list of factors found to be significant in the explanation of digital divide is mentioned in the following studies:

- ***Income and socio-economic status*** (Ebo, 1998; Chakraborty and Bosman, 2002; Pook and Pence, 2004; Barrantes and Galperin, 2008; Schleife, 2010);
- ***Occupation, skills and job experience*** (McLaren and Zappala, 2002; Rice and Katz, 2003; Losh, 2004; Tien and Fu, 2008; Wilbon, 2003);
- ***Gender and age*** (Trauth, 2002; DiMaggio et al., 2004; Noce and Mckeown, 2008; Abbey and Hyde, 2009; Orviska and Husdon, 2009);
- ***Education and literacy*** (Lim, 2002; Cornfield and Rainie, 2003; Peter and Valkenburg, 2006; Moon et al., 2010; Shirazi et al., 2010);
- ***Geographic location*** (Chen and Wellman, 2003; Akca, Sayili and Esengun, 2007; Prieger and Hu, 2008; Park and Jayakar, 2010);
- ***Ethnicity and race*** (Hoffman et al., 2000; Kim et al, 2007; Middleton and Chambers, 2009)
- ***Religiousness*** (Bell et al., 2004; Zilka, 2012)
- ***Proficiency of the English language*** (Foulger, 2001; Halpin et al., 2007; Alam et al., 2009)
- ***Family structure*** – number of children at home (Kennedy et al., 2003; Hitt and Tambe, 2007; Schleife, 2010)
- ***Speed and quality of internet service*** (Savage and Waldman, 2009; Glass and Stefanova, 2010).

Numerous studies conducted in the past decade have investigated the determinants for digital divide in Israel. Enoch and Soker (2006) studied the effects of social–structural factors on university students' use of web-based instruction. Their study used data from registration questionnaires of students at the Open University of Israel. In line with the results appearing in other digital divide studies, they have found that structural factors such as age, gender and ethnicity play a significant role in the continuous existence of the usage gap. Ganayem, Rafaeli and Azaiza (2009) analyzed the digital divide between Jews and Arabs and found considerable gaps between these two populations with regards to internet use. A later study conducted by Avidar (2009), suggests that the digital gap in Israel is diminishing due to greater diffusion of the internet nationwide. The access gap between Arabs and Jews seemed to decrease with age, as younger population better bridges the digital divide (Avidar, 2009).

Socio-demographic and Content Usage Attributes of the Digital Divide

According to Park (2009) and Van Deursen and Van Dijk (2014), as access to digital media and its use has become more common, the term has gained additional meanings and it could be used to designate effects other than access gaps. The concept of the digital divide has consequently shifted, so that it now refers to differences generated by the ***type of content*** different users are exposed to rather than whether they have access to information technologies.

The literature shows that the type of content people use differs by gender and age. Studies reveal that women, on the one hand, prefer religious content, health related information, online games and are more likely to use the Internet's communication tools. On the other hand, adult males are more likely to use the Internet for information, entertainment, commerce (Jackson et al., 2001; Subrahmanyam et al., 2001; Peter and Valkenburg, 2007; Park, Kim and Na, 2007; Zillien and Hargittai, 2009), online gaming (Schumacher and Morahan-Martin, 2001) and dating (Rudder, 2014).

Age appears to be one of the most significant variables that influence Internet use (Bonfadelli, 2002; Fox and Madden, 2005; Zillien and Hargittai, 2009). Studies show that young adults extensively use communication tools, such as instant messaging (IM) and chatting, and are more likely to pursue entertainment and leisure activities, such as gaming, downloading files or music (Howard et al., 2001; Dutton et al., 2011; Fox and Madden, 2005; Jones and Fox, 2009). In contrast, buying products online, e-mailing and searching for health-related information are more popular among older users (Jones and Fox, 2009).

Socio-economic status indicators were found to have a significant impact on Internet use (e.g. Zillien and Hargittai, 2009). DiMaggio et al. (2004) found that that persons of higher socio-economic status employ the Internet more productively and to greater economic gain than their less privileged, but nonetheless connected, peers. There is evidence to suggest that people with lower levels of socio-economic status tend to use the Internet in more general and superficial ways (Van Dijk, 2005).

A few studies suggest that education is the most important predictor for explaining the types of online activities a person will pursue (Robinson et al., 2003; Van Dijk, 2005). People with higher levels of education use the Internet for health information, financial transactions and research, while people with lower levels of education use the Internet for casual browsing, playing games or gambling online (Howard et al., 2001). Madden (2003) found that people with higher levels of education are less likely to download music or use instant messaging but are more likely to use the Internet for news, work, travel arrangement and product information. Hargittai and Hinnant (2008) found that those with higher levels of education use the Internet for ‘capital-enhancing’ activities, which include seeking political or government information, exploring career opportunities and consulting information about financial and health services. Helsper and Galacz (2009) revealed that the lower educated are least likely to use the Internet for educational and economic purposes, even when they have similar levels of Internet access and skills (Van Deursen and Van Dijk, 2014).

Although income is strongly correlated with education, some studies show an independent effect of income on physical and material Internet access (Katz and Rice, 2002). Concerning types of online activities, Madden (2003) revealed that higher income households, on the one hand, are less likely than low income households to use instant messaging or download music, but on the other hand, are more likely to seek news and product information, arrange for travel online and typically use the Internet for work (Van Deursen and Van Dijk, 2014).

The data in the above-mentioned studies were mostly collected through surveys and some by in-depth interviews and transaction data. Various statistical methods were employed in these studies to estimate the factors influencing digital divide, among them: regression (multiple regression, logit model and binomial-logistic regression); multivariate analysis (Mann-Whitney, Wilcoxon tests); structural equations modelling, continuous-time survival model, discrete choice model and tree-based technique.

The main drawback of domestic assessment studies, based on surveys and interviews, is that they are all prone to self-report bias. This may include one or more factors which may affect the reliability and validity of the research findings: honesty of response, introspective ability (the ability to provide an accurate response to the question), the degree of understating and interpretation of the question and difficulty in providing “accurate” measure in rating questions (Graham et al., 1993; Donaldson et al., 2002; Hoskin, 2012).

Using Digital Trace and “Big Data” for Measuring Digital Gaps

Monitoring and measuring the digital gap requires assessment of digital behavior by the use of new type of data and tools that go beyond the commonly used self-report measures and methodologies. The technological revolution witnessed in the past two decades, characterized by exponential computing growth, advancement in software, hardware, cloud and information technologies – has produced enormous opportunities, as well as challenges in the production and utilization of complex data. This can be especially observed in the context of **"digital trace data” and “Big Data”** framework and the development of various **unobtrusive research** tools to analyze them.

Unobtrusive research tools are used to analyze recorded human behavior. They can be collected without the subjects’ active involvement and they are especially valuable as they do not disturb or ‘break’ the data, therefore effectively capturing facts that circulate at a particular space and time (Webb 2000; Brabazon, 2010). One of the main strengths of unobtrusive research is the documentation of actual rather than self-reported behavior. Other advantages include repeatable results, easier access to data, continuity and the fact that permission for recording the data from subjects is usually not needed. Unobtrusive methods are relatively inexpensive and are appropriate for longitudinal studies that follow activities over a period of time (Kellehear,1993).

Digital trace or digital footprint data are “records of activity undertaken through online information systems. They are marks left as a sign of passage, a recorded evidence that something has occurred in the past” (Howison et al., 2011). Jones and Rafaeli (2000) used archaeology as an analogous field for describing the role of digital artefacts on society and human behavior: “Like archaeological tells, the remains of digital traces can supply evidence on human behavior and interaction”. O'Brien (2010), following Jones and Rafaeli (2000) has described the information age as an archeology site of modern existence. The definition of “Big Data” is complex and constantly changing. However, there is some consensus in the literature regarding its main characteristics, relating to three dimensions (Beyer and Laney, 2012): volume (vast data that cannot be handled by traditional analytical tools), velocity of production (the recording of real-time events) and variety (complex datasets including numerous sources of **digital traces or footprints**, such as unstructured text, images, videos and logs) and variety (digital traces relate to numerous types of records of activity undertaken through online information systems).

Within this context Callegaro and Yang (2018) outlined a typology of the main sources of digital traces and “Big Data”:

- Internet data - online text and multimedia: records of user behavior [browsing and search activity conducted by individuals in search engines (e.g. Google), E-Government and E-Commerce websites (e-shopping).
- Social media data (a specific subset of Internet data publicly available by mining social media networks, e.g. Twitter, Facebook, LinkedIn, YouTube, Instagram. It includes the analysis of various activities such as editing, reading and search.
- Website data and machine data (logs, traffic bandwidth, clickstream data, sensor readings, GPS system output, cookies, transactions, website analytics etc.).
- The Internet of Things data (traces from any device using the internet as communication transmission protocol).
- Behavioral data (a specific subset of the IOT based devices such as smartphones and wearables, recording locations or movements, various sensor data etc.).
- Transaction data (records of orders, shipments, payments, returns, billing, and credit card activities).
- Administrative data (national health records, taxes, benefits, pensions etc.) and commercial data (tracks from companies, businesses, consumers, users).

The analysis of these human and machine generated digital trace data sources, used in tandem with spatial and demographic layers, can provide us better understanding of the digital divide.

Visualizing the Digital Divide

Data visualizations are highly important for raising stakeholders’ interest and for strengthening the understanding and trust in the data (Cherchye et al. 2007). Design choices of visualization can influence the interpretation of various metrics and are therefore critical. Visualization is not a trivial issue (Nardo et al. 2005). Its complexity is derived from the data characteristics (as being hierarchical, multi-dimensional, time-oriented data), as well from its goals and tasks. Yet, there is lack in design guidelines for data visualization. Visualization of metrics based on "digital trace" data is challenging. For example, time intervals of different variables may not be the same. Another challenge might be items abstraction. Since "digital trace" data approach is innovative, variable visualization might help in the construction process of various metrics and indices.

Chapter 2: Methodology

In this project, we apply a novel approach to effectively measure and analyze the digital divide in Israel using human-generated, digital-trace data. An unobtrusive, triangulation-based approach is used to evaluate and analyze differences in online user-behavior between various socio-demographic groups.

Research Goal and Objectives

The main goal of the research is to supply the theoretical and practical underpinnings for measuring and evaluating the digital divide using digital trace data. The research objectives are as follows:

- To identify relevant “digital trace” variables which can be collected from human generated sources.
- To collect data samples of these digital trace artifacts.
- To process these digital trace records in a way that will facilitate the construction of meaningful, un-biased digital-divide measurements and indicators at a detailed spatial and sectoral (e.g. sociodemographic attributes) levels.
- To formulate a methodology for triangulating digital trace data.
- To triangulate various digital trace-based sources in order to deepen our understanding of the digital divide phenomenon and to construct more robust measurements, allowing evidence-based evaluation of actions.
- To deepen our research on the topic of indicator visualization, with focus on abstraction of "digital trace" data.
- To provide a proof of concept and set of insights regarding the use of digital trace analysis in the study of the digital divide and to its ability to replace self-report methods in this task.

The development of such research infrastructure could supply valuable inputs for policy evaluation and decision-making.

Research Motivation and Contribution

The literature review has highlighted the limitations of contemporary digital divide studies. These studies, using obtrusive techniques, are costly, usually one-shot and un-replicable, provide little continuous or even benchmark measures and are subject to real sampling and self-report biases. In contrast to these techniques, unobtrusive methods such as digital trace methods are becoming more relevant than ever, providing useful tools such as ‘data-mining’ and cultural analytics to better understand the huge amount of data surrounding us and the evolution of social behavior and communication on a digital platform (O'Brien, 2010).

The current project offers several novel methodological, theoretical and practical contributions to the study of digital divide. First, to the best of our knowledge, no research to this date has offered a comprehensive methodological framework for measuring and analyzing the socio-demographic aspects of the digital divide within a country using multi-source digital trace data. Thus, laying the foundations and techniques for the construction of indicators using digital trace data constitutes a clear and significant methodological contribution to the body of knowledge.

Second, there is a genuine theoretical contribution in the development of indicators for the evaluation of the digital divide phenomenon. These indicators are continuous and unobtrusive measures. They are “living indicators” - easy to produce and replicate. Due to their on-demand availability, low cost, virtually unlimited and constantly updated number of observations (population instead of sample), these unobtrusive digital trace measures could provide a powerful, dynamic and spatially detailed account of the spatial divide problem. This could facilitate a better understanding and enable a more refined chronological and spatial measuring of the digital divide.

Third and lastly, the project offers several practical novelties, crucial for the work of policy and decision makers. The ability of digital trace techniques to capture data “on demand”, to facilitate benchmarking and to present relevant indicators over time and space at a detailed level, provides decision makers and stakeholders the ability to receive high resolution data at the sectoral and levels. This type of high-end resolution is missing in the National ICT index, despite being an extremely groundbreaking enterprise due to its use of advanced techniques. The indicator visualizations also contribute in raising stakeholder’s interest, promoting transparency, understanding and trust in the data, and are of high relevance to the use of the media and the public at large.

Research Questions

On the basis of the goals and objectives presented above, five sets of research questions were developed:

- Which socio-demographic and locational factors (e.g. as age, gender, income, education and geographical location) best explain on-line user behavior?
- Do significant digital gaps exist between different socio-demographic groups in terms of on-line user behavior (e.g. in terms of volume and variety of website visits)?
- How do these digital gaps, parsed with socio-demographic factors, are reflected in terms of the internet content consumed. What kind of patterns can be observed?
- In what way can the triangulation of various sources of digital trace data be used to deepen and broaden our understanding of the digital-gap phenomena?
- What kind of tailored methodological tools could be built for triangulating digital trace data and how these tools could be mobilized for the study of digital gaps?

Research Population and Data

The research population is composed of Israeli on-line internet users. In the framework of the research, six different digital trace data sources were used to analyze online user behaviour, with the specific aim of studying digital gaps. Four datasets were based on aggregated user data (SimilarWeb online, SimilarWeb Learning Set, Google Trends, Google Analytics) and two datasets were based on individual/user level data (Ifat Panel Data and Buzzilla). All data sources are based on digital traces and reflect actual or revealed user behaviour. The following paragraphs present a short description of each source:

- **Ifat Panel Data**: The dataset is owned by the Ifat Media Advertising Monitoring Company, a subsidiary of the Ifat Group. It is based on a panel of 993 on-line users, comprising a sample of the Israeli population. The panel was originally selected for the purpose of tracking advertisement clicks, in selected, leading Israeli websites (e.g. YouTube, Ynet, Mako). By the request of SNI, the tracking software embedded in the users' web-browsers was adjusted to record all on-line activity and website visits conducted between October 15th, 2017 and November 15th, 2017. The digital trace data recorded in this dataset includes the full URL list and the date and time of entrance to each website. The on-line activity of each unique, anonymous user (URL visits) is matched with his or her socio-demographic attributes such as gender, age, income, education, geographical location and religiousness level. A Taxonomy of the full URLs into content usage categories (e.g. e-shopping, e-learning, finance, search, social networks, leisure, entertainment, e-gov etc.) has been performed by SNI researchers (see the categorization protocol in Annex 1). The categorization process was conducted as follows: Out of two million website entries, conducted by the panel of 993 on-line users, 41,518 unique URLs were identified. Each unique record was manually coded to represent a single major category. Overall, 31 website categories were coded. In some cases, sub-categories were coded as well (e.g. “Institutions” in “e-Health” and in “education”). The use of string functions assisted in the identification of websites belonging to the same category (e.g. “bank”, “pay” and other strings for the Finance category; “doctor”, “health”, “clalit” and other strings for the e-Health category etc.). Overall, 90.45% of activity was coded, representing 18.3% of all websites. About 9.7% of the activity was categorized as “junk”, as it represented panel-oriented activity (e.g. connecting to various panel websites).

- **SimilarWeb On-line platform**: A digital platform based on data extracted from four main sources: 1. A panel of web surfers made of millions of anonymous users equipped with a portfolio of apps, browser plugins, desktop extensions and software. 2. Global and Local Internet Service Providers. 3. Web traffic directly measured from a learning set of selected websites and apps intended for specialized estimation algorithms. 4. A colony of web crawlers that scan the entire Web and apps stores. SimilarWeb collects anonymous clickstream data from a diverse panel of users and employs algorithms to estimate overall metrics for web and apps. Available metrics include: total visits, traffic share (desktop, mobile), global and country rank, average visit duration, pages per visit, bounce rate, traffic share by country and region, visits by gender and by age groups etc. The platform, including various web tools, covers the last 24 month period of the on-line activity (SimilarWeb, 2016).

- **SimilarWeb Learning Set Database**: an aggregated data subset extracted from the on-line platform. The database includes four types of variable categories: **socio-demographic variables** (gender, age-group, geographic region, language), **user-behaviour variables** (website URL, website category, website subcategory, device used), **temporal variable** (month) and **website visits and unique user summary variables** (total number of visits, total number of unique users in each observation). The dataset is aggregated in its nature and includes 178,094 data points, representing all possible combinations of the variables present in the first three variable category groups. It is important to note that the SimilarWeb Learning Set covers only a very small fraction of the total online activity of Israeli users (only 80 websites are covered).

The database includes digital trace data on more than 48 million website visits for the year 2017.

- **Buzzilla:** A digital platform for monitoring and tracking social media and information from forums, groups and message boards, collecting millions of responses (talkbacks) to articles, forum posts, and blogs in various fields. This data pool is used for conducting social media research on themes such as conversation topics. The platform allows to perform segmentation of communities and participants and to measure the volume of activity. Our examination covered social media discussions conducted between October 15th, 2017 and November 15th, 2017.

- **Google Trends:** An online search tool that allows the user to see how often specific keywords, subjects and phrases have been queried over a specific period of time. This tool works by analyzing a portion of Google searches to compute how many searches have been done for the terms entered, relative to the total number of searches conducted on Google over the same time. The service provides information on the search query volumes of its users since January 2004 and allows researchers to select searches by geographical region (provinces, states, countries), categories and sub-categories (e.g., travel, finance, food), and frequency (daily, weekly, monthly). Results are displayed in a graph that Google calls "Search Volume Index". The data in the graph can be exported to a csv file and edited in Excel or other spreadsheet applications (Siliverstovs and Wochner, 2018).

- **Google Analytics:** An online analytics service that provides webmasters with a wide variety of information about the activity that takes place on their website. Google Analytics enables website owners to segment their visitors, study traffic trends and optimize conversion funnels. Data is viewed by metrics which measure behavior and by dimensions which describe who the customers are. Metrics and dimensions help website owners to answer fundamental questions such as who visits their websites and what are they doing. Common data that can be analyzed in GA include: aggregate page views, total number of visitors, number of unique visitors, website visiting time, geographic location of visitors (on a country, state and city level), specific terms that users searched, etc.[3].

[3] https://www.bigcommerce.com

Table 1 below, presents a summary of the data sources used in the framework of this research by their main characteristics.

Table 1: Description of data sources by main characteristics

Data source	**Ifat Panel**	**SimilarWeb On-line**	**SimilarWeb Learning Set**	**Buzzilla**	**Google Trends**	**Google Analytics**
Type	Dataset	Tool and dataset	Dataset	Tool and dataset	Tool and dataset	Tool and dataset
Level of data aggregation	anonymized user level	Aggregated	Aggregated	anonymized user level	Aggregated	Aggregated
Time period (project)	October-November 2017	October-November 2017	January-December 2017/Oct-Nov 2017	October-November 2017	October-November 2017	October-November 2017
Socio-demographic dimension	Gender, age, income, education, geographical location and religiousness level	Gender, age, geographical location	Gender, age, geographical location	None	Language and geographical location	Gender, age, geographical location
Device coverage	Desktop	Desktop	Desktop and mobile	Desktop and mobile	Desktop and mobile	Desktop and mobile
Type of access	Specially tailored for the project	Subscription needed	Available for research purposes.	Subscription needed	Free access	Administrator permission required

Research Workplan and Methods

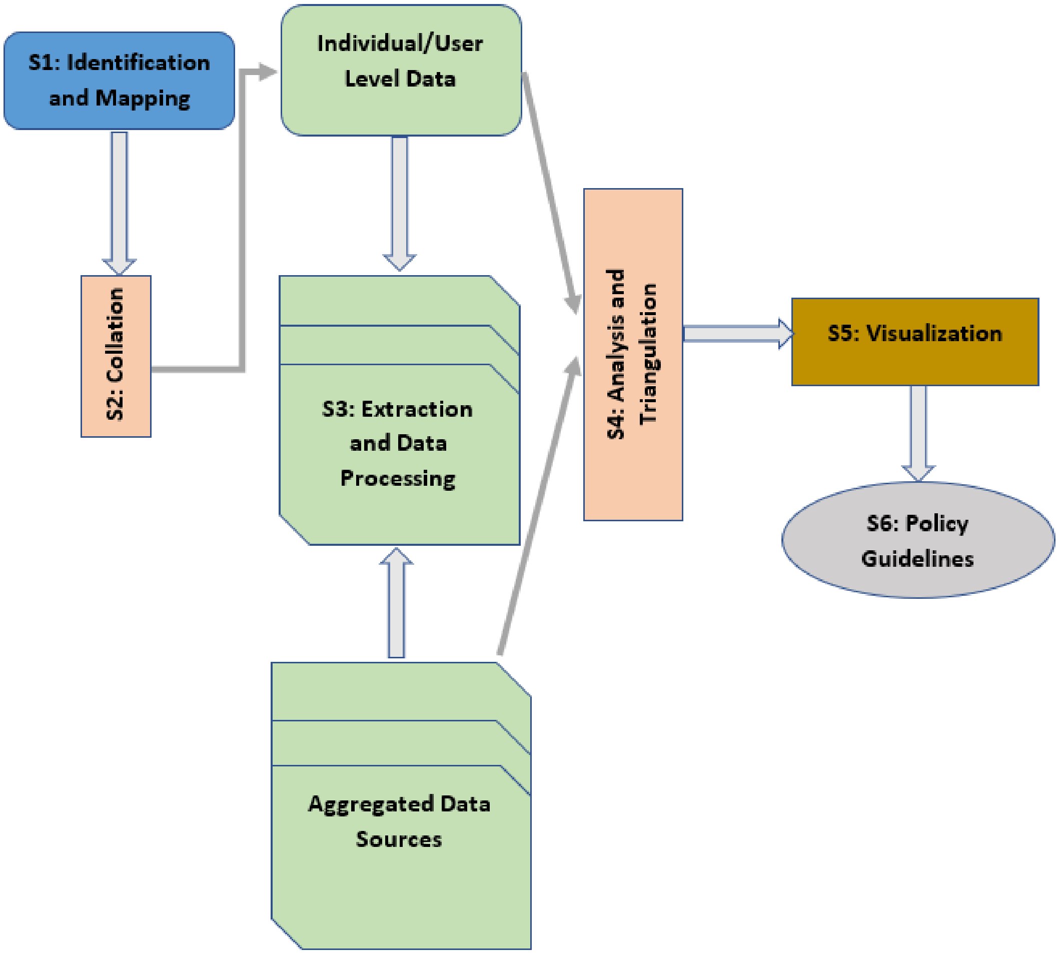
The research focuses on identifying, collecting, analyzing and visualizing unobtrusive, digital traces data that reflect on digital gaps. Figure 1 presents the research work plan, describing the various stages of the project.

- **Stage 1 (Identification and Mapping)** - included surveying publicly available or obtainable data sources which reflect digital gaps. Six various data sources were identified in the process, as described in detail in the section above.
- **Stage 2 (Collation)** - involved triaging data sources according to API nature, geo-tagging and timestamping ability, etc.
- **Stage 3 (Extraction and Data Processing)** - the relevant data sources were either downloaded, extracted (Google Analytics, Google Trends, Buzzilla, SimilarWeb On-line platform) or supplied to SNI as raw datasets (Ifat Panel Data, SimilarWeb Learning Set). The raw datasets (Ifat Panel Data and the SimilarWeb Learning Set) were then cleaned and processed (e.g. deletion of missing or un-valuable data, coding of

variables and variable values, aggregation of URLs into content usage categories and sub-categories, development of new variables etc.).

- **Stage 4 (Analysis and Triangulation**) - the various data sources were analyzed on a separate as well as on a triangulation basis in order to create both a detailed and a “bird-eye” view of the digital divide phenomenon. The triangulation process is exemplified by a case study on rights realization and is presented in Chapter 4.
- **Stage 5 (Visualization)** - involved the presentation of the data via visualization techniques.
- **Stage 6 (Policy Guidelines)** - policy guidelines and insights were formulated for the benefit of various stakeholders and the research community.

Figure 1: Research work plan


S1: Identification and Mapping
Individual/User Level Data
S2: Collation
S3: Extraction and Data Processing
S4: Analysis and Triangulation
S5: Visualization
S6: Policy Guidelines
Aggregated Data Sources

Chapter 3: Analyzing the Digital Divide in Israel: What do digital traces tell?

In this chapter, we apply a novel approach to effectively map overt online-user behavior by the analysis of digital trace data. Previous studies have investigated the socio-economic and content usage aspects of the digital divide using data extracted from questionnaires, interviews and surveys. In this particular study, unobtrusive digital trace data from various human sources, parsed with reference to socio-demographic and spatial attributes (e.g. age, gender, income, education and geographical location) are used to study and analyze digital gaps in Israel. The research employs a wide range of descriptive and qualitative research methods and tools, including graphs, two-dimensional tables, statistical tests (t-test, ANOVA, Post-hoc tests), regression models and a specially tailored normalized index, in order to map and analyze these digital traces.

In the analysis process, two main sources of trace data are used: the Ifat Panel Dataset and the SimilarWeb Learning Set. The "digital gaps" in these two sources are defined as differences in terms of ***usage volume*** (number of visits/distribution of visits), differences in ***variety*** (the number of different website categories visited by the user) and the differences in the ***content usage*** (e.g. the type of on-line activities or content consumed, defined by the volume of visits per content usage category). The Ifat Panel Dataset accounts for on-line activity conducted between October 15th, 2017 and November 15th, 2017. It covers the activity of 993 unique users (visitors) in about 1.6. million entries (visits) in about 41,500 websites. The SimilarWeb Learning Set, covers on-line activity conducted in 2017. It covers the activity unique users (visitors) and web entries (visits) in 80 popular websites.

Socio-demographic Aspects of the Digital Divide in Israel

Digital Gaps Reflected by Usage Volume

Table 2 presents descriptive statistics for website visits, parsed by socio-demographic and geographical attributes - gender, age, language and geographical location. Tests for differences in mean visits (t-test or ANOVA) are also included in the table. As can be observed from the data, male users exhibit higher usage volume than female users. The mean website visits for male users was found to be 33% higher than that of female users. The difference in means was found to be significant at the 0.001 level. As can be observed from the data, website visits decrease with age and significant gaps in usage volume exist between the age groups (P<0.001). The 25-35 age group is the most active in terms of website visits, exhibiting 2.5 times higher usage volume than the 65+ group. Post-hoc tests for differences between pair of means (LSD tests) for the six age groups show that the 25-35 age group statistically differs (P<0.001) from all other age groups, exhibiting higher mean differences (Table 3). The 65+ age group also differs (P<0.001) from all other age groups with respect to mean visits, exhibiting lower mean differences.

Table 2: Descriptive statistics for usage volume (visits), parsed by socio-demographic and geographical attributes and tests for means differences (t-test/ANOVA) in website visits

		VISITS			
		N	Total visits	Mean	Standard Deviation
Gender*	Female	84735	19769973	233.3	1092.3
	Male	93357	28950569	310.1	1497.5
Age*	18-24	29322	8480994	289.2	1146.0
	25-34	38835	15900929	409.4	2126.1
	35-44	31822	8462003	265.9	1120.9
	45-54	28644	6297804	219.9	960.5
	55-64	28792	6194035	215.1	902.4
	65+	20677	3384777	163.7	522.9
Language*	Arabic	1697	257794	151.9	390.1
	Hebrew	109437	35367448	323.2	1588.3
	Russian	16202	2177388	134.4	292.5
Region*	Jerusalem District	18311	2213016	120.9	340.8
	North District	15604	1464939	93.9	231.6
	Haifa District	20288	2086510	102.8	264.7
	Tel Aviv District	74060	36131593	487.9	1982.1
	Center District	33214	5300955	159.6	509.1
	South District	16137	1465106	90.8	224.8
	*. The mean difference is significant at the 0.001 level (for t-tests or ANOVA)				

Source: Special SNI data processing of the SimilarWeb Learning Set

Significant differences (P<0.001) in usage volume can also be observed between Hebrew, Arabic and Russian speakers. Here, it is important to note that the language of the user is proxied by the user-selectable setting of the web browser, generally defaulting to the language of the operating system. The usage volume among Hebrew speakers is two times larger than Arabic speakers and 2.4 times larger than native Russian speakers.

Table 3: Post-hoc tests (LSD) between age groups, accounting for differences in pair of means (visits)

(I) Age	(J) Age	Mean Difference (I-J)	N	Std. Error
18-24 (n=29322)	25-34	-120.212*	38835	10.200
	35-44	23.320*	31822	10.673
	45-54	69.372*	28644	10.953
	55-64	74.106*	28792	10.939
	65+	125.539*	20677	11.973
25-34 (n=38835)	18-24	120.212*	29322	10.200
	35-44	143.532*	31822	9.969
	45-54	189.584*	28644	10.269
	55-64	194.318*	28792	10.253
	65+	245.751*	20677	11.350
35-44 (n=31822)	18-24	-23.320*	29322	10.673
	25-34	-143.532*	38835	9.969
	45-54	46.052*	28644	10.738
	55-64	50.786*	28792	10.724
	65+	102.219*	20677	11.777
45-54 (n=28644)	18-24	-69.372*	29322	10.953
	25-34	-189.584*	38835	10.269
	35-44	-46.052*	31822	10.738
	55-64	4.734	28792	11.003
	65+	56.167*	20677	12.031
55-64 (n=28792)	18-24	-74.106*	29322	10.939
	25-34	-194.318*	38835	10.253
	35-44	-50.786*	31822	10.724
	45-54	-4.734	28644	11.003
	65+	51.433*	20677	12.018
65+ (n=20677)	18-24	-125.539*	29322	11.973
	25-34	-245.751*	38835	11.350
	35-44	-102.219*	31822	11.777
	45-54	-56.167*	28644	12.031
	55-64	-51.433*	28792	12.018
	*. The mean difference is significant at the 0.05 level			

Source: Special SNI data processing of the SimilarWeb Learning Set

Post-hoc tests (Table 4) conducted for analyzing differences between pair of means for three language groups show that statistically significant differences exist in the usage volume between Hebrew speakers and Arabic speakers (P<0.001) and between Hebrew speakers and Russian speakers (P<0.001), but not between Russian and Arabic speakers.

Table 4: Post-hoc tests (LSD) between language speakers, accounting for differences in pair of means (visits)

(I) Language	(J) Language	Mean Difference (I-J)	N	Std. Error
Arabic	Hebrew	-171.265*	109437	32.286
(n=1697)	Russian	17.522	16202	33.675
Hebrew	Arabic	171.265*	1697	32.286
(n=109437)	Russian	188.786*	16202	11.110
Russian	Arabic	-17.522	1697	33.675
(n=16202)	Hebrew	-188.786*	109437	11.110
	*. The mean difference is significant at the 0.05 level			

Source: Special SNI data processing of the SimilarWeb Learning Set

Table 5: Post-hoc tests (LSD) between geographical regions, accounting for differences in pair of means (visits)

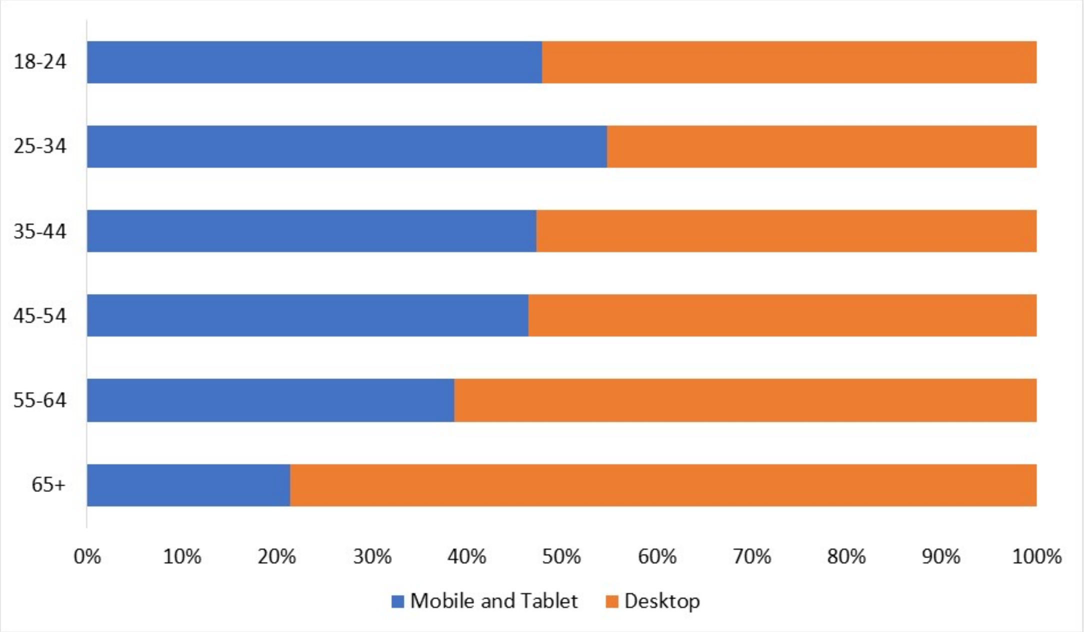
(I) Region	**(J) Region**	**Mean Difference (I-J)**	**N**	**Std. Error**
Jerusalem District	North District	26.975	15604	14.272
	Haifa District	18.013	20288	13.353
(n=18311)	Tel Aviv District	-367.012***	74060	10.811
	Center District	-38.743**	33214	12.057
	South District	30.065*	16137	14.144
North District	Jerusalem District	-26.975	18311	14.272
	Haifa District	-8.962	20288	13.948
(n=15604)	Tel Aviv District	-393.987***	74060	11.539
	Center District	-65.718***	33214	12.714
	South District	3.091	16137	14.707
Haifa District	Jerusalem District	-18.013	18311	13.353
	North District	8.962	15604	13.948
(n=20288)	Tel Aviv District	-385.025***	74060	10.380
	Center District	-56.755***	33214	11.672
	South District	12.053	16137	13.817
Tel Aviv District	Jerusalem District	367.012***	18311	10.811
	North District	393.987***	15604	11.539
(n=74060)	Haifa District	385.025***	20288	10.380
	Center District	328.269***	33214	8.651
	South District	397.077**	16137	11.380
Center District	Jerusalem District	38.743**	18311	12.057
	North District	65.718*	15604	12.714
(n=33214)	Haifa District	56.755***	20288	11.672
	Tel Aviv District	-328.269***	74060	8.651
	South District	68.808***	16137	12.570
South District	Jerusalem District	-30.065*	18311	14.144
	North District	-3.091	15604	14.707
(n=16137)	Haifa District	-12.053	20288	13.817
	Tel Aviv District	-397.077***	74060	11.380
	Center District	-68.808***	33214	12.570
	*. The mean difference is significant at the 0.05 level. **. The mean difference is significant at the 0.01 level. ***. The mean difference is significant at the 0.001 level.			

Source: Special SNI data processing of the SimilarWeb Learning Set

The data presented in Table 2 shows stark gaps and statistically significant differences between various geographic regions (P<0.001) with respect to visits. The usage volume of internet users from the Core region (Tel Aviv district) is five times larger than those of users from the country's periphery (South and North Districts). Post-hoc tests for differences between pair of means (Table 5) for six geographic regions show that the Tel Aviv district statistically differs (P<0.001) from all other geographic districts, exhibiting higher mean visits.

Figure 2 and Figure 3 present the usage volume distribution by age and device type and by gender and device type. As can be seen from , the usage volume distribution of older age groups (65+, 55-64) in mobile devices significantly differs from the distribution of the other age groups. Only 21% of the on-line activity of the 65+ age cohort and 38% of the activity of the 55-64 age cohort is carried out in mobile devices (smartphones and tablets), whereas this activity is substantially higher in the younger age cohorts (comprises 46%-55% of the total usage volume).

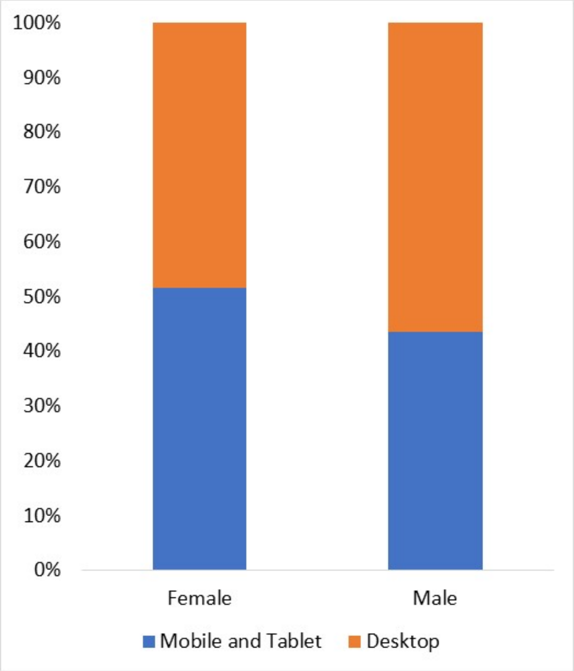
Figure 2: Usage volume distribution (visits), breakdown by age and device type



Source: Special SNI data processing of the SimilarWeb Learning Set

No substantial differences in this respect can be identified between the genders (Figure 3), whereas the on-line activity of female users (in terms of website visits volume) in mobile devices is only slightly higher than those of male users.

Figure 3: Usage volume distribution (visits), breakdown by gender ande device type



Source: Special SNI data processing of the SimilarWeb Learning Set

Modeling the Relationship between Socio-demographic Attributes and Usage Volume

A multiple log-linear regression approach for modelling the relationship between the socio-demographic variables and usage volume (visits) was applied on the SimilarWeb Learning Set data and is presented in Table 6. The independent variables include: age (Age); A dummy variable representing female users (Female_D); A set of dummy variables representing the user's language: Arabic_D, Russian_D, Hebrew_D (the reference variable excluded from the model is all other languages); A set of dummy variables denoting the geographical location of the user: Core_D (composed of the Tel Aviv and central districts) and Periphery_D (composed of the North and South Districts). The reference dummy variable excluded are the Haifa and Jerusalem districts.

As can be seen from the table, all variables are statistically significant at the 0.001 level. The user's age is negatively correlated with the number of visits. The older the users are, the lower their number of visits in on-line websites. Female users are negatively correlated with the number of visits, implying higher share of internet usage by male users. A statistically significant relationship exists between the language or the ethnical background of the user and the number of visits. Russian and especially Arab speakers are negatively correlated with the number of visits on the one hand, and Hebrew speakers are positively correlated with the number of visits, on the other hand. The strongest and most statistically significant predictor of website visits (t=94; b=0.732) is geographical location at the Core (Tel Aviv and Central Districts). This finding implies that geographical location contributes the most to digital divides, as seen by the clear spatial dichotomy between the Core and the Periphery with respect to usage intensity.

Table 6: Regression model explaining usage volume

Model		Unstandardized Coefficients		Standardized Coefficients	t
		B	Std. Error	Beta	
1	(Constant)	3.978*	.011		375.059
	Age	-.052*	.002	-.062	-27.026
	Female_D	-.214*	.006	-.079	-34.669
	Arabic_D	-.444*	.032	-.032	-13.805
	Russian_D	-.127*	.012	-.027	-10.674
	Hebrew_D	.206*	.007	.074	29.337
	Core_D	.732*	.008	.263	94.321
	Periphery_D	-.104*	.010	-.029	-10.581
Dependent Var: LN_VISITS. R=0.297; N=177,437; * : B is significant at the 0.001 level					

Source: Special SNI data processing of the SimilarWeb Learning Set

Digital Gaps Reflected by Internet Content Diversity

Internet content diversity is an indicator marking the range of interest in terms of internet content consumption. The average number of website categories visited by a unique user is used as a proxy for estimating diversity. This metric can be used to evaluate users' behavior and to identify and analyze digital gaps between various socio-economic groups. In order to evaluate the content diversity, we use the Ifat Panel Data which includes 31 different content or "activity" categories (see Annex 2).

Table 7 provides descriptive statistics for website diversity, parsed by socio-demographic (gender, income level, education level and religiousness level) and geographical (region) attributes of the unique users. As can be seen from the table, male users are more diverse than female users with respect to internet content consumption, visiting on average 18.1 different internet content categories as compared to 16.3 categories visited by females. These differences were found to be statistically significant at the 0.001 level. The diversity level seems to slightly rise with age. The diversity level of the 55+ age group was found to be significantly higher (P<0.001) than those of all other age groups with the exception of the 45-54 age group (no significant differences).

Table 7: Descriptive statistics for website diversity, parsed by socio-demographic and geographical attributes

		Diversity (website categories)		
		Count	Standard Deviation	Mean
Gender	Female	615	5.6	16.3
	Male	373	5.4	18.1
Age	15-17	53	6.1	16.0
	18-24	146	5.8	15.5
	25-34	329	5.8	16.5
	35-44	240	5.2	17.7
	45-54	124	5.1	18.2
	55+	96	5.2	18.3
Income level	Significantly less than average	141	5.9	16.6
	Slightly less than average	167	5.7	16.7
	Average	176	5.5	16.5
	Slightly above average	229	5.7	17.2
	Significantly above average	75	5.8	16.3
Education Level	Primary or less	7	4.0	20.6
	Secondary without matriculation	109	6.2	16.7
	Secondary with matriculation	203	5.7	15.8
	Post-secondary non academic	207	5.7	17.1
	Bachelor's degree	322	5.4	17.4
	Master’s degree or higher	135	5.4	17.8
Religiousness level	Secular	569	5.5	17.2
	Traditional (Shomrei Masoret)	208	5.9	16.2
	Religious	145	5.7	17.1
	Ultra-Orthodox	66	5.6	17.8
Geographic region	Jerusalem Metro	109	6.0	17.4
	Tel Aviv Metro	325	5.6	17.3
	Haifa Metro and North	238	5.7	16.8
	South and Shefela	216	5.5	16.7
	Sharon	100	5.3	16.6

Source: Special SNI data processing of Ifat Panel data

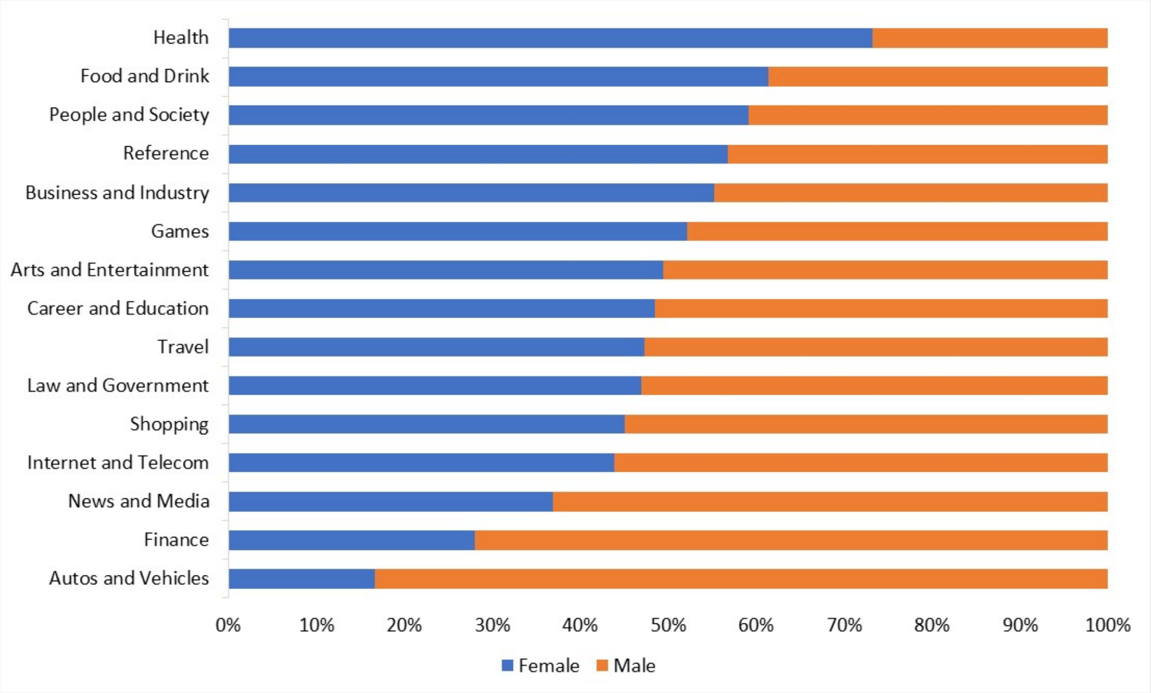
The data shows that relatively small differences in the diversity level exist with respect to income level. The level of diversity seems to rise with the education level, whereas individuals with post-secondary education or higher have a substantially higher diversity level (P<0.001) than individuals holding secondary education or lower (with the exception of the primary or less group, which is very small). Surprisingly, the ultra-orthodox group exhibits the highest level of diversity and statistically differs from all other groups (P<0.001). Spatial differences with respect to the diversity level can be also observed from Table 7. Users from the Tel Aviv metropolitan region and the Jerusalem metropolitan region exhibit the highest diversity levels and statistically differ from users from other regions (P<0.001).

Digital Gaps Reflected by Differences in Content Usage

As discussed in detail in the previous chapter, the scientific literature provides considerable evidence for the existence of disparities between various socio-economic groups with regards to the type of internet content consumed by users (e.g. news, gaming, e-health). However, these findings were mostly based on self-report, obtained from obtrusive methods (e.g. surveys). In order to evaluate the differences in content usage between various groups using digital trace data, we employ the website taxonomy and categorization methodology described in Chapter 2. As mentioned in the previous chapter, the website categorization of the Ifat Panel Data, encompassing 31 different website categories (see Annex 2), was conducted by SNI researchers based on their subjective evaluation, whereas the SimilarWeb taxonomy (including 15 categories and 30 sub-categories) was pre-defined and embedded within the SimilarWeb Learning Set Database (see Annex 3 and Annex 4).

Figure 4 describes the content usage distribution by gender for the SimilarWeb database, broken down by category. The metric expresses the distribution of website visits in each on-line activity/content category. As can be seen from the figure, e-health activities are dominated by females, accounting for 73% of the total visits volume in this category. Additional content usage category dominated by women is “Food and Drink”, accounting for 61% of total visits. As can be seen from Table 8, exhibiting t-test for equality of mean visits by categories, significant differences exist between female and male users with respect to mean visits in the Health category (P<0.001) and in the Food and Drink category (P<0.01). As can be also observed, some content usage categories are exclusively dominated by male users. These include Autos and Vehicles (accounting for 83% of total visits), Finance (accounting for 72% of total visits) and News and Media (accounting for 63% of total visits). The differences in mean visits between the genders are significant at the 0.001 level for all these three categories (Table 8). The data also shows that the content usage among females and males is much more evenly distributed in Gaming, Arts and Entertainment and Career and Education domains (Figure 4), with no significant differences between the genders (Table 8).

Figure 4: Content usage distribution by gender, breakdown by category (SimilarWeb)



Source: Special SNI data processing of the SimilarWeb Learning Set

Table 8: Differences in content usage, breakdown by gender and category (SimilarWeb)

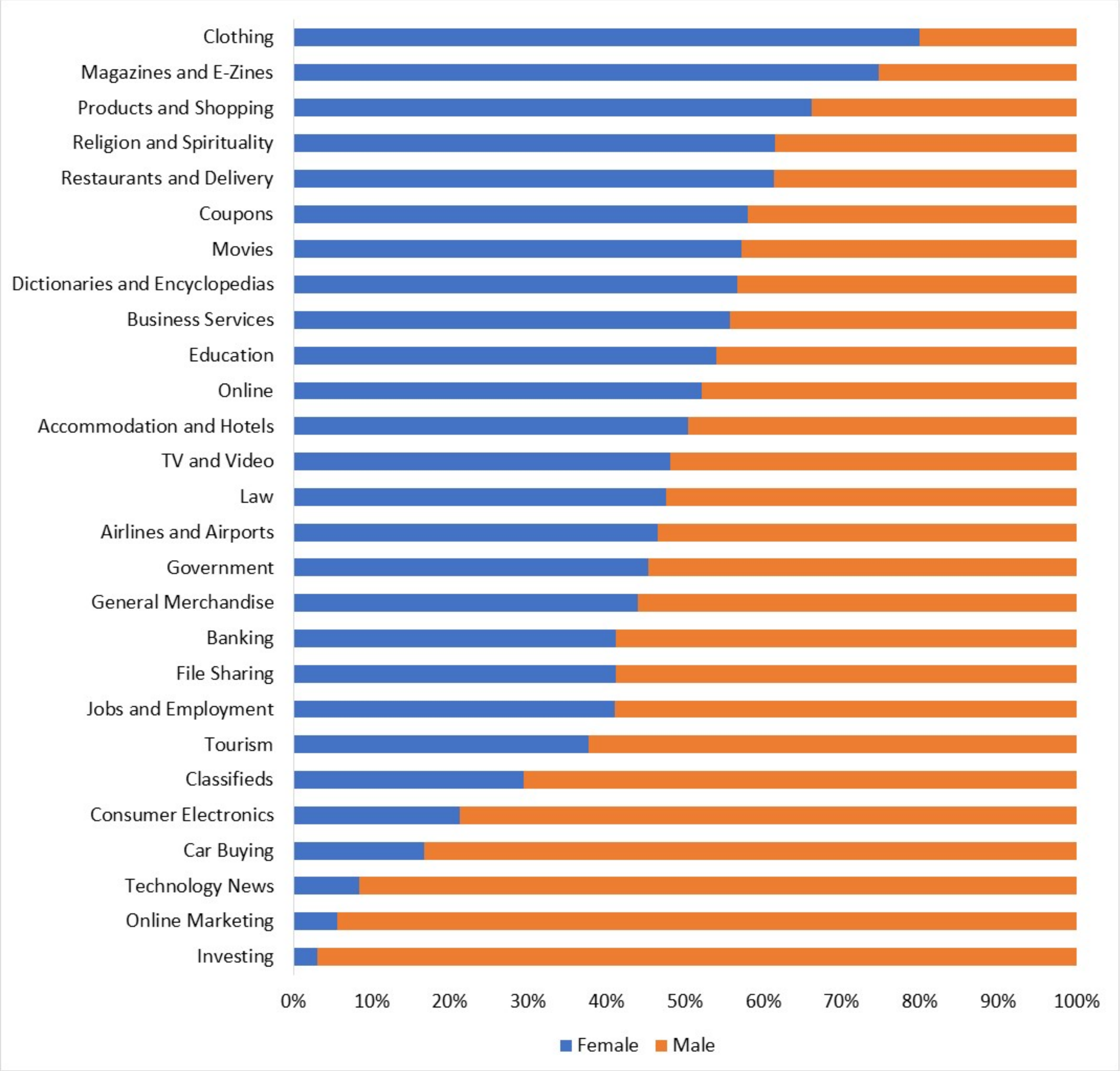
	VISITS	VISITS	VISITS		VISITS	
			Gender		Gender	
			Female	Male	Female	Male
Category	Sum	Column N %	Row N %	Row N %	Mean	Mean
Internet and Telecom	**946238**	**2.8%**	**46.6%**	**53.4%**	**180.8***	**201.5***
Arts and Entertainment	2964760	9.4%	50.2%	49.8%	177.3	182.9
News and Media	**33729636**	**45.3%**	**44.8%**	**55.2%**	**349.5*****	**486.9*****
Shopping	**4023242**	**13.6%**	**48.6%**	**51.4%**	**155.9*****	**180.6*****
Business and Industry	**1280896**	**3.8%**	**50.9%**	**49.1%**	**209.9***	**176.7***
Autos and Vehicles	**64881**	**0.7%**	**26.7%**	**73.3%**	**32.2*****	**58.7*****
Travel	961275	3.9%	49.2%	50.8%	135.5	146.2
Health	**934290**	**4.8%**	**61.2%**	**38.8%**	**131.7*****	**76.2*****
People and Society	**677092**	**3.4%**	**52.9%**	**47.1%**	**128.1*****	**99.4*****
Law and Government	883673	3.1%	47.3%	52.7%	160.5	163.1
Finance	**883744**	**2.9%**	**39.9%**	**60.1%**	**123.3*****	**210.6*****
Career and Education	302338	2.2%	47.5%	52.5%	81.6	78.6
Games	122527	0.6%	50.9%	49.1%	124.6	118.4
Reference	**394391**	**1.9%**	**51.7%**	**48.3%**	**127.9***	**104.7***
Food and Drink	**343977**	**1.6%**	**55.4%**	**44.6%**	**133.8****	**104.7****

*. The mean difference (t-test) is significant at the 0.05 level.
**. The mean difference (t-test) is significant at the 0.01 level.
***. The mean difference (t-test) is significant at the 0.001 level.

Source: Special SNI data processing of the SimilarWeb Learning Set

Some very interesting insights on gender related divides are obtained by a more fine-grained analysis (sub-category). Figure 5 describes the content usage distribution by gender for the SimilarWeb database, broken down by sub-category. The metric expresses the distribution of website visits in each sub-category. As can be seen from the figure, this high-resolution analysis sharpens, for example, the differences between the genders in on-line shopping behavior (which is not as stark when looking at the category breakdown in Figure 4), whereas female users completely dominate the Clothing sub-category (accounting for 80% of total visits) on the one hand and male users dominate the Consumer Electronics sub-category (accounting for 79% of total visits) on the other hand.

Figure 5: Content usage distribution by gender, breakdown by sub-category



Source: Special SNI data processing of the SimilarWeb Learning Set

As can be seen from Table 9, exhibiting t-test for differences in mean visits by sub-category, significant differences exist between female and male users in the Clothing sub-category (P<0.001) and in the Consumer Electronics sub-category (P<0.001).

Table 9: Differences in content usage (sub-categories), breakdown by gender and sub-category (SimilarWeb)

		Visits	Visits		Visits	
			Gender		Gender	
	Sum	Column N %	Female Row N %	Male Row N %	Female Mean	Male Mean
Online Marketing	**87540**	**0.4%**	**13.7%**	**86.3%**	**86***	**231***
TV and Video	1636631	10.8%	49.2%	50.8%	148	155
General Merchandise	**1675157**	**1.7%**	**38.0%**	**62.0%**	**1134*****	**885*****
Car Buying	**64881**	**1.3%**	**26.7%**	**73.3%**	**32*****	**59*****
Business News	**2978221**	**13.2%**	**40.9%**	**59.1%**	**121*****	**297*****
Accommodation and Hotels	289664	1.6%	49.2%	50.8%	186	177
Magazines and E-Zines	**52603**	**0.1%**	**62.1%**	**37.9%**	**666****	**369****
Clothing	**141489**	**2.1%**	**70.0%**	**30.0%**	**75*****	**44*****
Movies	**189187**	**1.9%**	**52.6%**	**47.4%**	**106***	**88***
Business Services	**1164536**	**5.3%**	**50.9%**	**49.1%**	**241***	**199***
File Sharing	**111985**	**0.6%**	**45.2%**	**54.8%**	**165***	**194***
Airlines and Airports	**628407**	**4.9%**	**49.5%**	**50.5%**	**120***	**135***
Tourism	**43204**	**0.3%**	**45.4%**	**54.6%**	**118*****	**162*****
Government	245578	3.0%	46.6%	53.4%	79	83
Coupons	**401178**	**3.2%**	**52.5%**	**47.5%**	**137*****	**109*****
Newspapers	**4783024**	**12.8%**	**46.3%**	**53.7%**	**313*****	**426*****
Religion and Spirituality	**551732**	**4.8%**	**54.2%**	**45.8%**	**129*****	**96*****
Technology News	**140342**	**1.2%**	**22.5%**	**77.5%**	**45*****	**139*****
Products and Shopping	**74331**	**1.5%**	**58.6%**	**41.4%**	**57*****	**41*****
Investing	**305991**	**1.0%**	**14.4%**	**85.6%**	**64*****	**341*****
Jobs and Employment	130107	2.0%	42.1%	57.9%	64	67
Online	122527	1.0%	50.9%	49.1%	125	118
Consumer Electronics	**313779**	**2.9%**	**29.4%**	**70.6%**	**77*****	**119*****
Dictionaries and Encyclopedias	**394391**	**3.4%**	**51.7%**	**48.3%**	**128****	**105****
Banking	**577753**	**4.0%**	**46.3%**	**53.7%**	**128*****	**158*****
Education	172231	1.8%	53.5%	46.5%	96	94
Law	638095	2.4%	48.2%	51.8%	259	266
Restaurants and Delivery	**343977**	**2.8%**	**55.4%**	**44.6%**	**134****	**105****
Sports News	**4249382**	**3.9%**	**23.5%**	**76.5%**	**222*****	**1368*****
Classifieds	**576988**	**4.1%**	**43.0%**	**57.0%**	**96*****	**174*****

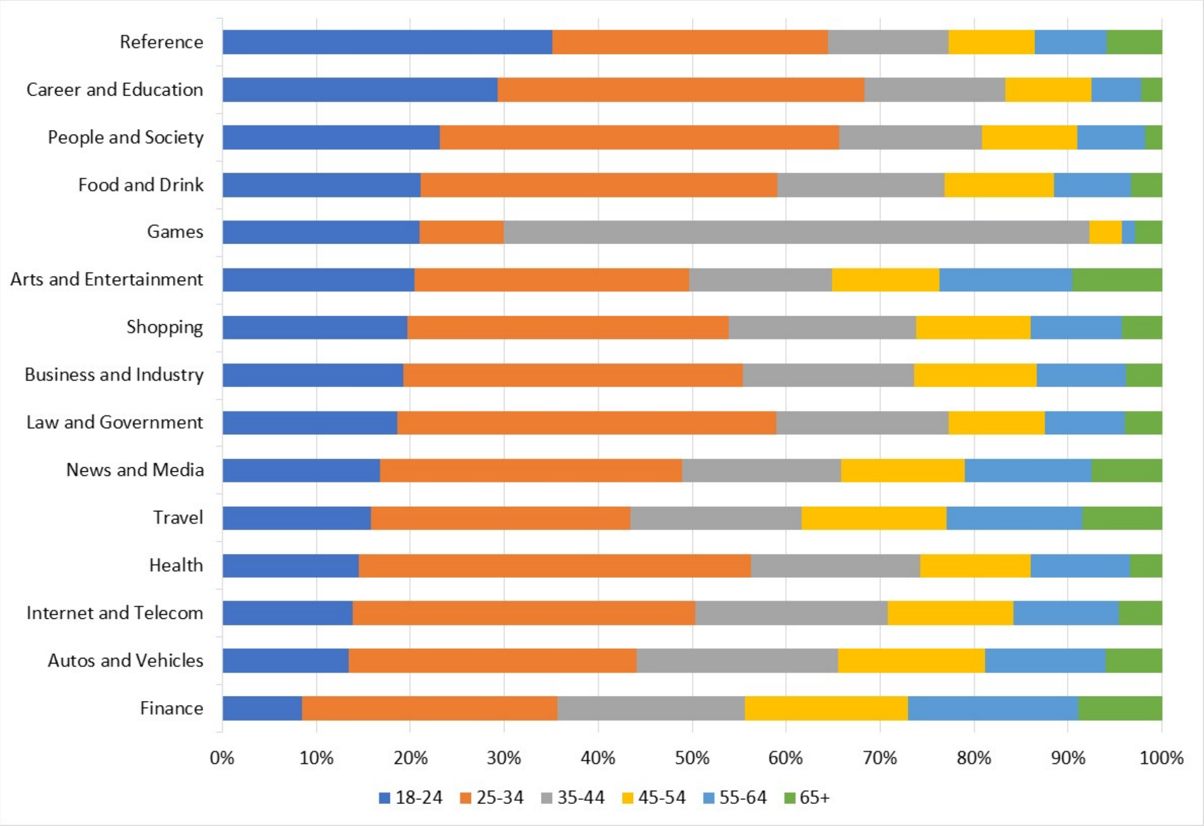
Source: Special SNI data processing of the SimilarWeb Learning Set

*. The mean difference (t-test) is significant at the 0.05 level.
**. The mean difference (t-test) is significant at the 0.01 level.
***. The mean difference (t-test) is significant at the 0.001 level.

Additional “masculine” and “feminine” on-line behavior can also be identified. For example, the “Investing” and “Technology News” sub-categories are totally dominated by male users (accounting for 97% and 92% of total visit volume respectively), whereas the “Magazines and E-Magazines” and “Religion and Spirituality” sub-categories are occupied by female users (accounting for 75% and 62% of total visit volume respectively). The differences in mean visits between the genders are significant at the 0.01 level for all these four sub-categories (Table 9). It is important to note that for some on-line activities, no significant differences between males and females were identified in terms of usage volume or mean visits (e.g. consumption of TV and Video content, E-education and E-government services, making online reservations for hotels and other tourism related accommodations, jobs and employment search etc.).

The distribution of usage activity by age groups (Figure 6) sheds an interesting light on digital divides:

Figure 6: Content usage distribution by age, breakdown by category (SimilarWeb)



Source: Special SNI data processing of the SimilarWeb Learning Set

As can be seen from Figure 6, the on-line behavior of younger age cohorts (18-24, 25-34, 35-44) differs from the behavior of older age cohorts (45-54, 55-64, 65+) with respect to the type and volume of the content consumed. Younger age cohorts are characterized by their high-share of usage of gaming activities, career and education related activities, law and online government services (e-gov), health services (on-line health websites) and the use of reference services (e.g. online dictionaries). In contrast, older age cohorts are characterized by high volume usage of Finance, Travel (e.g. online plane tickets and hotel reservations) and arts and entertainment activities.

In contrast to the SimilarWeb Learning Set Database which focuses on the online activity of millions of unique users in a relatively limited number of websites, the Ifat Panel data is based on a group of 993 on-line users and tracks their on-line activity in a very large number of websites. Due to the relatively small sample size of the panel, it is vital to use a normalized index for the description of content usage, especially when parsing it with socio-demographic attributes, in order to prevent a biased representation. The general form for this index is given by the following equation:

$$\mathrm{RCU} = \left[\frac{\mathrm{V}(C_t S_x)}{V(C_t S \sum_{x=1}^{n})}\right] \Big/ \left[\frac{V(C \sum_{t=1}^{n} S_X)}{V(C \sum_{t=1}^{n} S \sum_{x=1}^{n})}\right]$$

Whereas:

RCU= Relative Content Usage Index

V=volume (in terms of category visits)

C= Internet content usage (on-line activity) category

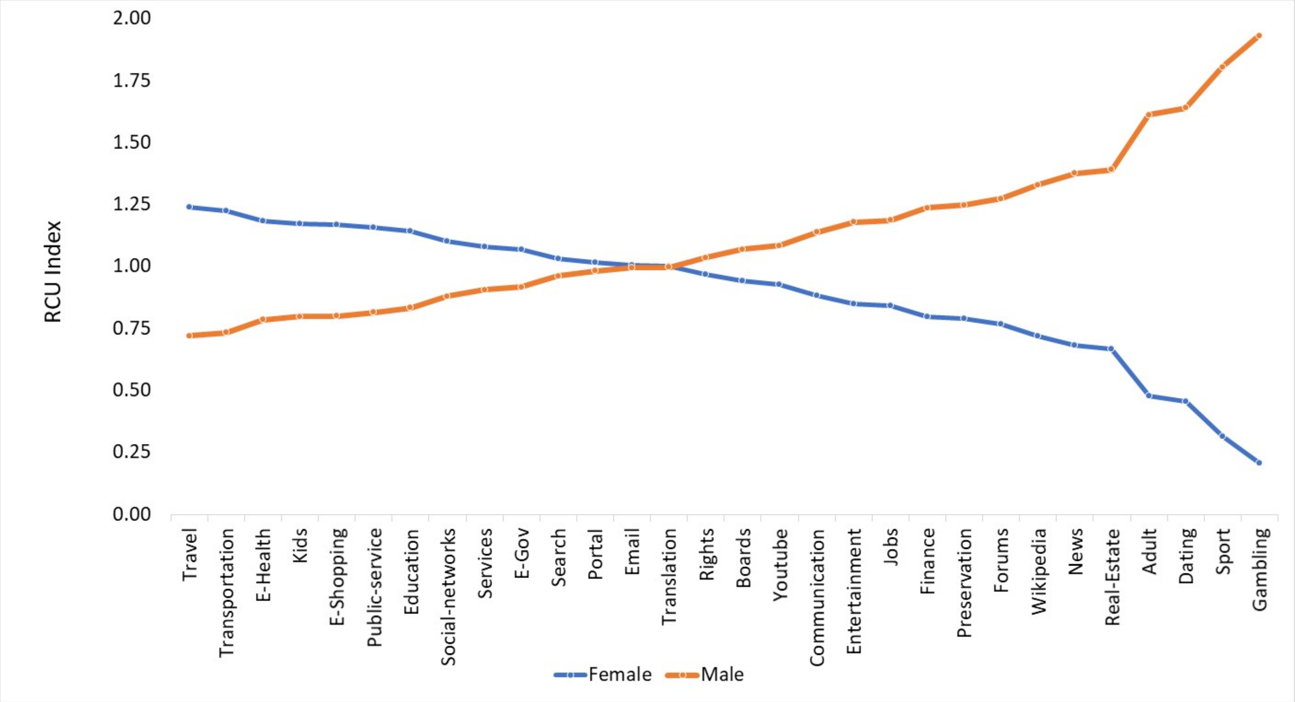
S=Socio-economic attribute

In the particular analysis presented below, we use two subgroups (e.g. for gender: male and female; for income level: low to medium income and high income etc.) for each socio-economic attribute (X=1,….2) and 31 categories (Search, E-mail, News, Finance etc.) for the content usage (t=1,…..31).

The Relative Content Usage Index enables to conduct comparisons between socio-demographic groups across the internet content categories. A value of “1” denotes “neutrality” or a lack of dominance of a particular segment (e.g. male or female) in a given content usage category (e.g. e-health, e-gov, etc.), whereas a value above ‘1’ denotes dominant or intensive usage of a particular segment in a particular content category (“high volume usage”) and a value below “1” marks low volume of usage in a given content category by a particular subgroup.

Figure 7 describes gender differences in the *Relative Content Usage Index* (RCU) across 31 content usage or online activity categories.

Figure 7: RCU Index parsed by gender and content usage category (Ifat Panel)



As can be seen from the left-hand side of the figure, the RCU among female users is substantially higher than that of male users in several activities. These include travel (1.7 times higher), Transportation (1.7 times higher), E-health (1.5 times higher), Kids (1.5 times higher) and E-Shopping (1.5 times higher). The right-hand side of the “scissor effect” describes digital gaps between male and female users. Here too, large gaps between male and female users in the RCU Index can be observed in several categories. Some notable examples include Gambling (9.3 times higher), Sports (5.7 times higher), Dating (3.6 times), Adult (3.4 times higher), Real Estate (2.1 times higher), News (2.0 times higher), Wikipedia (1.8 times higher), Forums (1.7 times higher) and Finance (1.6 times higher). The intersection point, marking RCU values of 1 or near 1 mark on-line activates where no substantial differences between the genders can be observed. These include: Search, Portal, E-mail, Translation, Rights realization and Boards.

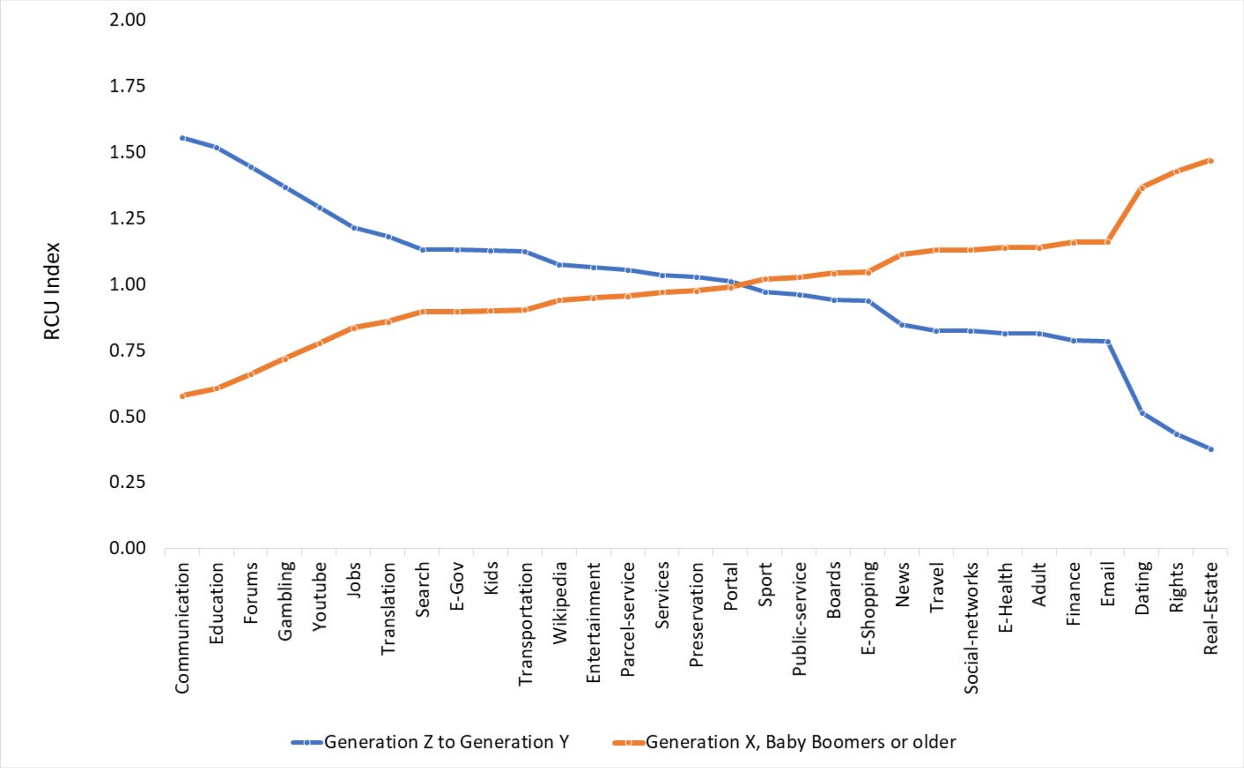
Table 10: Differences in content usage, breakdown by gender (Ifat Panel)

	Gender					
	Female			**Male**		
	Count	**Standard Deviation**	**Mean**	**Count**	**Standard Deviation**	**Mean**
Boards	615	72.9	21.5	373	136.4	34.4
Communication	615	23.8	5.7	373	45.4	10.3
Dating	**615**	**46.8**	**5.7***	**373**	**201.3**	**28.6***
E-Gov	615	59.7	24.9	373	78.2	30.0
E-Health	615	54.6	22.9	373	54.8	21.3
E-Shopping	615	749.2	297.8	373	650.5	286.0
Education	615	166.1	64.8	373	198.3	66.2
Email	**615**	**208.8**	**69.4***	**373**	**183.3**	**96.6***
Entertainment	**615**	**153.3**	**64.5***	**373**	**510.0**	**125.8***
Finance	**615**	**96.2**	**56.1****	**373**	**387.8**	**121.9****
Forums	**615**	**67.2**	**12.1***	**373**	**118.6**	**28.2***
Gambling	**615**	**3.5**	**0.5***	**373**	**49.1**	**6.5***
Jobs	615	96.2	18.5	373	164.3	36.6
Kids	615	32.0	5.6	373	27.8	5.4
News	**615**	**111.3**	**39.0***	**373**	**410.8**	**110.6***
Parcel-service	615	19.2	2.8	373	11.9	1.7
Adult	**615**	**47.2**	**4.3****	**373**	**88.1**	**20.1****
Portal	615	69.5	29.2	373	105.2	39.6
Preservation	**615**	**5.1**	**1.1***	**373**	**10.7**	**2.3***
Public-service	615	29.6	8.8	373	27.4	8.7
Real-Estate	615	8.9	1.3	373	51.7	3.8
Rights	614	9.1	3.0	373	14.7	4.5
Search	**615**	**272.8**	**204.8****	**373**	**299.5**	**268.1****
Services	615	86.7	28.0	373	55.3	32.9
Social-networks	615	204.3	92.4	373	197.9	103.3
Sport	**615**	**33.7**	**5.2*****	**373**	**120.4**	**41.4*****
Translation	615	19.1	5.7	373	36.5	7.9
Transportation	615	31.9	7.2	373	12.2	6.0
Travel	615	106.0	25.0	373	56.3	20.4
Wikipedia	**615**	**38.9**	**7.0***	**373**	**89.0**	**18.2***
YouTube	**615**	**34.3**	**19.7***	**373**	**91.0**	**32.3***
Source: SNI data processing of Ifat Panel Dataset	*. The mean difference (t-test) is significant at the 0.05 level. **. The mean difference (t-test) is significant at the 0.01 level. ***. The mean difference (t-test) is significant at the 0.001 level.					

Table 10 describes gender differences in content usage for 31 different activity categories for the Ifat Panel Data. The summary statistics included in the table are sample size, the number of mean visits and standard deviation for each category, parsed by gender. As can be seen from the table, the mean number of visits of male users is much higher than the average number of visits by female users in almost every category. Significant differences (higher means for male users) were found in the following categories: Dating, Email, Entertainment, Finance (e.g. banks, pension, online payments etc.), Forums, Gambling, News, Adult, Search, Sports, Wikipedia and YouTube. Higher mean visits values for female users could be found in only two categories (E-shopping, E-health), however these were not found to be statistically significant.

Figure 8 describes differences in the RCU Index between two aggregated age cohorts representing “young” (Generation Z to Generation Y – ages 18 to 34; n=528) and “older” (Generation X, Baby Boomers or older - age 35+; n=460) users.

Figure 8: RCU Index parsed by age group and content usage category (Ifat Panel)



As can be seen from the left-hand side of the figure, the RCU Index among Generation Z to Generation Y users is substantially higher than that of Generation X, Baby Boomers or older users in several activities. These include Communications (2.7 times higher), Education (2.5 times higher), Forums (2.2 times higher), Gambling (1.9 times), YouTube (1.7 times higher) and Jobs (1.5 times higher). The right-hand side of this “scissor effect” figure highlights the activities which are dominated by the Generation X, Baby Boomers or older generation users. Some notable examples include Real-Estate (3.9 times higher), Rights Realization (3.3 times higher), Dating (2.7 times higher), E-mail (1.5 times higher) and Finance (1.5 times higher). No substantial differences between these two age groups can be observed in the following activities: Wikipedia, Entertainment, Service, Preservation, Portals, Sports, Public-service, Boards and E-Shopping.

Table 11 describes differences in content usage parsed by age cohorts. As can be seen from the table, the mean number of visits by Generation X, Baby Boomers or older users is much higher than that of Generation Z to Generation Y users in almost every category. Significant differences (higher means for Generation X, Baby Boomers or older users) were found in the following categories: Boards (P<0.05), Dating (P<0.05), E-Health (P<0.001), E-Shopping (P<0.01), Email (P<0.001), Finance (P<0.001), News (P<0.01), Portal (P<0.05), Public-service (P<0.05), Search (P<0.05), Services (P<0.05), Social-networks (P<0.001) and Travel (P<0.01). Higher mean visit values for Generation Z to Generation Y users were found in only two categories – Communication (P<0.05) and Education (P<0.01).

Figure 9 describes differences in the RCU Index between by education level. As can be seen from the left-hand side of the figure, the RCU Index of users holding post-secondary education or lower level education is substantially higher that of users holding a bachelor’s degree or higher in several activities. These include Forums (2.4 times higher), Gambling (2.3 times), Kids (1.8 times higher), E-mail (1.5 times higher) and YouTube (1.4 times higher). The right-hand side of this “scissor effect” figure highlights the activities which are dominated by users holding a bachelor’s degree or higher level education. These include Real-Estate (3.6 times higher), Rights Realization (3.2 times higher) and News (1.3 times higher). No substantial differences between the two education groups can be observed in the following activities: Transportation, Dating, Search, Wikipedia, E-Shopping, Entertainment, Jobs and E-Health.

Table 11: Differences in content usage, breakdown by age groups (Ifat Panel)

	Age Generation					
	Generation Z to Generation Y			**Generation X, Baby boomers or older**		
	Count	**Standard Deviation**	**Mean**	**Count**	**Standard Deviation**	**Mean**
Boards	**528**	**68.68**	**20.04***	**460**	**129.39**	**33.68***
Communication	**528**	**44.23**	**9.28***	**460**	**13.63**	**5.23***
Dating	**528**	**50.58**	**5.94***	**460**	**181.49**	**23.92***
E-Gov	528	66.44	24.55	460	68.28	29.48
E-Health	**528**	**34.59**	**14.66*****	**460**	**69.99**	**31.07*****
E-Shopping	**528**	**455.41**	**221.91****	**460**	**918.21**	**375.41****
Education	**528**	**175.77**	**80.01****	**460**	**181.03**	**48.45****
Email	**528**	**86.50**	**50.46*****	**460**	**274.24**	**113.21*****
Entertainment	528	166.24	75.42	460	460.47	101.69
Finance	**528**	**117.41**	**51.41*****	**460**	**344.24**	**114.83*****
Forums	528	105.16	21.21	460	69.66	14.69
Gambling	528	35.41	3.05	460	23.47	2.43
Jobs	528	145.01	24.84	460	101.32	25.86
Kids	528	33.62	5.05	460	26.49	6.11
News	**528**	**278.83**	**45.14****	**460**	**256.05**	**90.03****
Parcel-service	528	15.56	2.04	460	18.13	2.80
Adult	528	37.79	6.72	460	87.89	14.25
Portal	**528**	**85.93**	**27.04****	**460**	**83.22**	**40.12****
Preservation	528	6.46	1.28	460	8.89	1.83
Public-service	**528**	**18.35**	**6.78***	**460**	**37.17**	**11.00***
Real-Estate	528	7.30	0.69	460	47.03	4.07
Rights	528	9.99	3.00	460	231.39	15.00
Search	**528**	**237.51**	**209.22***	**460**	**329.43**	**251.14***
Services	**528**	**48.09**	**24.95***	**460**	**99.12**	**35.47***
Social-networks	**528**	**132.18**	**64.24*****	**460**	**254.92**	**133.56*****
Sport	528	79.92	14.78	460	80.97	23.54
Translation	528	19.39	6.22	460	33.73	6.85
Transportation	528	21.72	6.10	460	30.60	7.44
Travel	**528**	**58.55**	**15.48****	**460**	**116.24**	**32.17****
Wikipedia	528	33.16	9.78	460	85.04	12.97
Youtube	528	75.74	25.52	460	42.13	23.30
Source: SNI data processing of Ifat Panel Dataset	*. The mean difference (t-test) is significant at the 0.05 level. **. The mean difference (t-test) is significant at the 0.01 level. ***. The mean difference (t-test) is significant at the 0.001 level.					

Figure 9: RCU Index parsed by education level and content usage category (Ifat Panel)

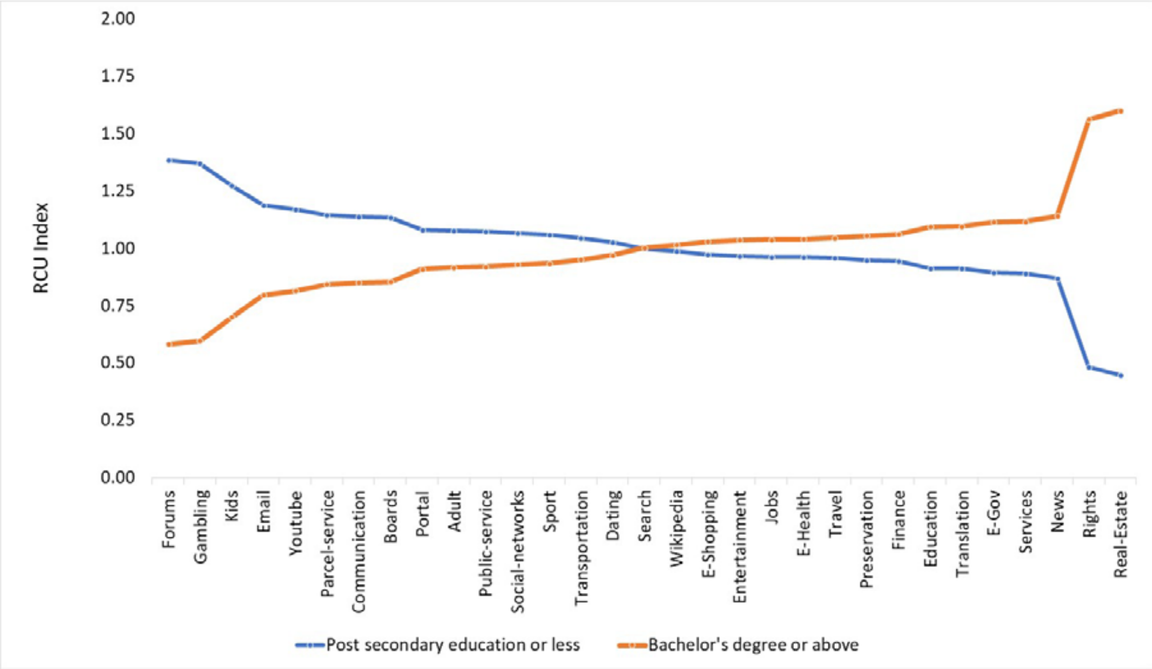


Table 12 describes education differences in content usage across the 31 activity categories. As can be seen from the table, significant differences in mean visits could be found in only two categories – Forums (P<0.05) and E-mail (P<0.05) activities. In these two content categories, users holding post-secondary education or lower level education exhibited statistically higher means than users holding a bachelor's degree or higher level education.

Table 12: Differences in content usage, breakdown by education level (Ifat Panel)

	EDUCATION LEVEL					
	Post-secondary education or less			**Bachelor's degree or above**		
	Count	**Standard Deviation**	**Mean**	**Count**	**Standard Deviation**	**Mean**
Boards	526	122.22	29.11	457	71.95	23.26
Communication	526	38.27	8.22	457	27.71	6.53
Dating	526	149.00	14.36	457	103.56	14.42
E-Gov	526	57.11	23.40	457	77.52	30.94
E-Health	526	55.00	20.84	457	54.08	23.87
E-Shopping	526	634.71	278.67	457	797.47	312.82
Education	526	151.25	58.29	457	206.70	73.97
Email	**526**	**255.72**	**92.33***	**457**	**103.69**	**65.74***
Entertainment	526	212.97	82.61	457	439.77	94.06
Finance	526	212.75	74.59	457	291.74	88.95
Forums	**526**	**118.74**	**24.59***	**457**	**36.77**	**10.98***
Gambling	526	35.72	3.71	457	23.08	1.71
Jobs	526	117.46	23.85	457	136.88	27.28
Kids	526	37.50	6.88	457	19.72	4.02
News	526	285.51	56.16	457	250.57	77.98
Parcel-service	526	18.56	2.68	457	14.64	2.10
Adult	526	70.19	10.77	457	61.50	9.71
Portal	526	90.28	35.05	457	78.71	31.26
Preservation	526	8.32	1.43	457	6.95	1.68
Public-service	526	33.15	9.17	457	22.87	8.34
Real-Estate	526	8.29	0.99	457	47.02	3.76
Rights	526	13.24	4.05	457	232.02	13.90
Search	526	260.04	222.08	457	311.47	236.83
Services	526	52.15	25.95	457	97.19	34.51
Social-networks	526	224.32	99.81	457	172.40	92.35
Sport	526	89.01	19.54	457	69.97	18.28
Translation	526	17.54	5.80	457	34.98	7.40
Transportation	526	29.15	6.83	457	22.57	6.59
Travel	526	85.25	21.63	457	96.49	25.02
Wikipedia	526	68.90	10.81	457	55.58	11.80
YouTube	526	77.98	27.89	457	37.22	20.60
Source: SNI data processing of Ifat Panel Dataset	*. The mean difference (t-test) is significant at the 0.05 level.					

Figure 10 describes differences in the RCU Index between income levels. As can be seen from the left-hand side of the figure, users characterized by low to medium income levels exhibit substantial higher RCU Index values than users characterized by high-income levels in several activities. These include Rights realization (5.1 times higher), Dating (3.0 times), Jobs (2.5 times higher), Portal (1.9 times higher), Gambling (1.6 times higher), Boards (1.6 times higher) and YouTube (1. 5 times higher). The right-hand side of this "scissor effect" figure highlights the activities which are dominated by users characterized by high-income levels. These include Real-Estate (4.7 times higher), Finance (1.8 times higher), Travel (1.7 times higher), News (1.6 times higher) and Services (1.5 times higher). No substantial differences between the genders can be observed in the following activities: Kids, Search, Preservation, Wikipedia, Entertainment, Email, E-Shopping, E-Health.

Figure 10: RCU Index parsed by income level and content usage category (Ifat Panel)

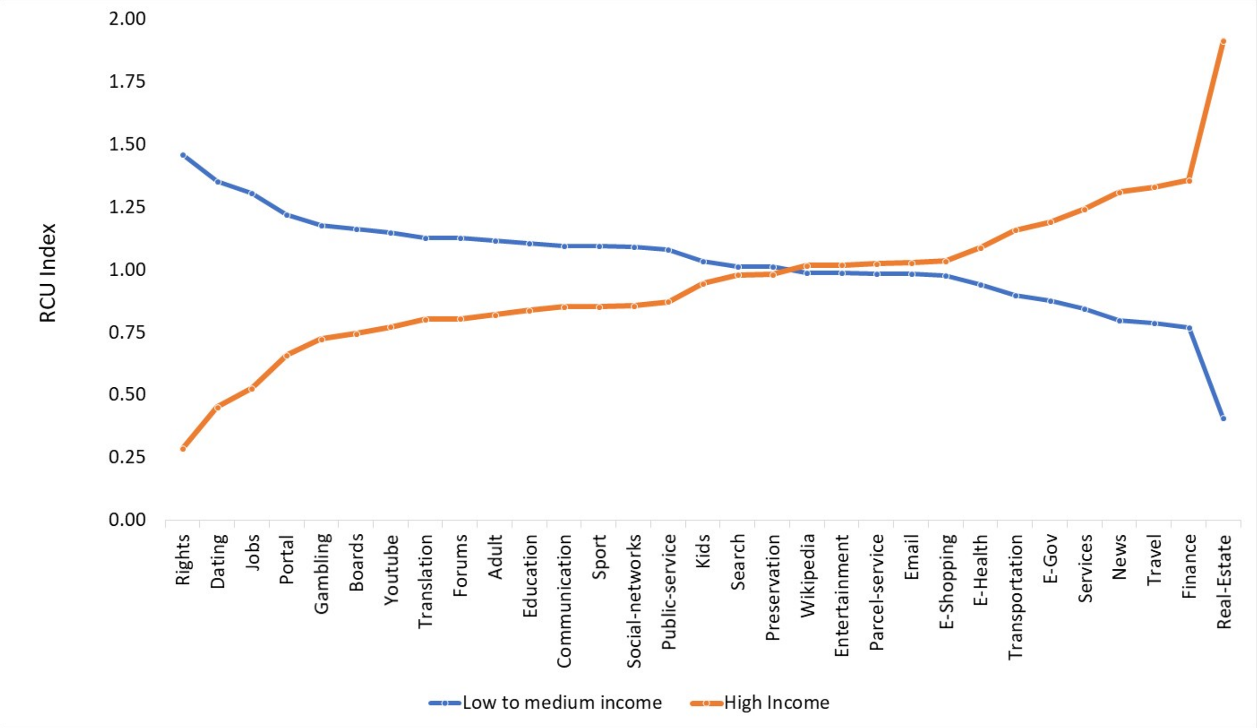


Table 13: Differences in content usage, breakdown by income level (Ifat Panel)

	Income level					
	Low to medium income			High Income		
	Count	Standard Deviation	Mean	Count	Standard Deviation	Mean
Boards	484	114.65	28.08	304	57.49	18.50
Communication	484	39.95	7.89	304	23.24	6.33
Dating	484	64.86	8.88	304	29.26	3.05
E-Gov	484	52.08	23.68	304	91.97	33.01
E-Health	484	60.15	21.06	304	54.63	25.01
E-Shopping	484	825.64	284.94	304	581.27	310.35
Education	484	171.82	75.11	304	202.59	58.65
Email	484	114.43	74.14	304	294.02	79.66
Entertainment	484	165.74	71.99	304	283.08	76.33
Finance	**484**	**145.55**	**59.88***	**304**	**396.10**	**108.63***
Forums	484	102.20	19.49	304	54.67	14.33
Gambling	484	25.11	2.46	304	12.70	1.55
Jobs	**484**	**154.37**	**33.12***	**304**	**65.81**	**13.76***
Kids	484	30.73	6.22	304	36.91	5.84
News	484	121.28	46.72	304	289.49	78.58
Parcel-service	484	18.23	2.36	304	16.16	2.52
Adult	484	59.79	9.86	304	64.77	7.46
Portal	**484**	**101.26**	**37.81****	**304**	**48.37**	**21.00****
Preservation	484	6.57	1.21	304	5.49	1.21
Public-service	484	29.23	8.72	304	12.72	7.23
Real-Estate	484	6.38	0.92	304	56.90	4.41
Rights	484	225.55	13.98	304	7.94	2.82
Search	484	271.60	223.58	304	323.16	222.40
Services	484	41.13	23.64	304	115.52	35.82
Social-networks	484	211.74	94.27	304	146.01	76.05
Sport	484	94.32	20.35	304	57.10	16.31
Translation	484	35.16	7.18	304	12.18	5.25
Transportation	484	22.79	5.87	304	35.24	7.78
Travel	484	66.09	17.43	304	123.32	30.27
Wikipedia	484	70.40	10.59	304	56.95	11.21
YouTube	**484**	**80.45**	**28.30***	**304**	**37.99**	**19.56***
Source: SNI data processing of Ifat Panel Dataset	*. The mean difference (t-test) is significant at the 0.05 level. **. The mean difference (t-test) is significant at the 0.01 level.					

Table 13 describes differences in mean visits between the two income levels across the 31 activity categories. As can be seen from the table, significant differences in mean visits could be found in only four categories. Users characterized by low to medium income levels exhibited statistically higher mean visits than users which are characterized by high-income levels in the Jobs (P<0.05), Portal (P<0.01) and YouTube (P<0.05) content usage categories. Users characterized by high income level exhibited higher mean visits than users which are characterized by low to medium income levels in the Finance (P<0.05) content category.

Figure 11 describes differences in the RCU Index by religiousness level:

Figure 11: RCU Index parsed by religiousness level and content usage category (Ifat Panel)

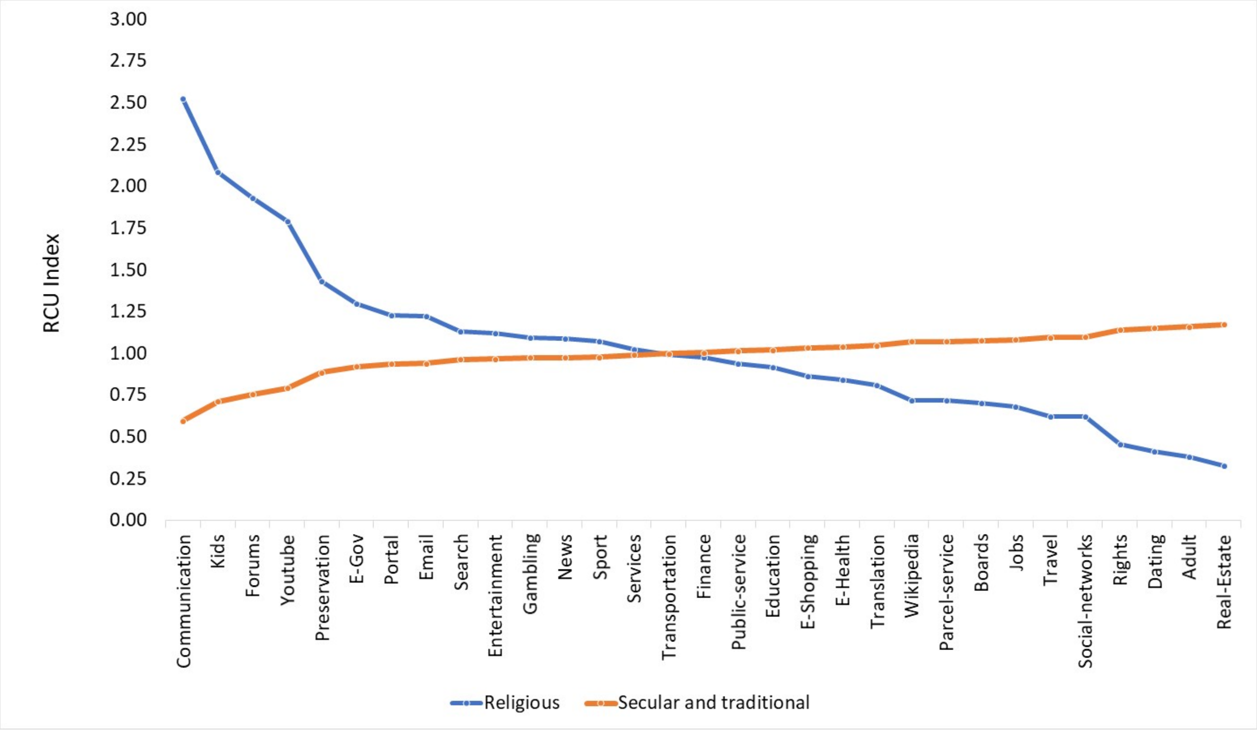


Table 14: Differences in content usage, breakdown by level of religiousness (Ifat)

	Level of religiousness					
	Secular and traditional			**Religious**		
	Count	**Standard Deviation**	**Mean**	**Count**	**Standard Deviation**	**Mean**
Boards	777	110.33	28.66	211	59.85	18.04
Communication	**777**	**14.94**	**4.47****	**211**	**66.04**	**18.17****
Dating	777	144.13	16.64	211	44.01	5.74
E-Gov	777	65.93	24.94	211	71.90	33.86
E-Health	777	58.40	23.38	211	37.47	18.33
E-Shopping	777	761.32	306.08	211	496.27	246.58
Education	777	189.35	67.19	211	133.35	58.41
Email	777	136.78	75.57	211	344.01	94.77
Entertainment	777	371.16	85.47	211	156.47	95.70
Finance	777	240.95	82.02	211	289.42	76.96
Forums	**777**	**71.98**	**13.84***	**211**	**137.60**	**34.11***
Gambling	777	29.60	2.72	211	33.38	2.94
Jobs	777	138.19	27.63	211	67.47	16.78
Kids	**777**	**22.64**	**3.99***	**211**	**49.35**	**11.26***
News	777	289.88	64.92	211	173.85	70.18
Parcel-service	777	17.16	2.59	211	15.44	1.67
Adult	**777**	**72.59**	**11.96***	**211**	**31.75**	**3.82***
Portal	777	76.10	31.36	211	111.41	39.67
Preservation	777	7.14	1.37	211	9.43	2.14
Public-service	777	30.15	8.95	211	22.93	7.97
Real-Estate	777	36.63	2.68	211	4.62	0.73
Rights	777	178.12	9.89	211	13.30	3.82
Search	777	284.35	222.43	211	285.24	251.95
Services	777	82.56	29.85	211	47.12	29.83
Social-networks	**777**	**217.82**	**106.84*****	**211**	**119.75**	**58.50*****
Sport	777	68.13	18.62	211	115.30	19.74
Translation	777	29.91	6.89	211	11.06	5.13
Transportation	777	28.18	6.79	211	17.24	6.50
Travel	**777**	**99.37**	**25.73***	**211**	**43.19**	**14.14***
Wikipedia	777	69.80	12.18	211	23.62	7.90
YouTube	777	32.48	19.57	211	118.17	42.62
Source: SNI data processing of Ifat Panel Dataset	*. The mean difference (t-test) is significant at the 0.05 level. **. The mean difference (t-test) is significant at the 0.01 level. ***. The mean difference (t-test) is significant at the 0.001 level.					

As can be seen from the left-hand side of Figure 11, religious users (ultra-orthodox and national religious) exhibit substantially higher RCU values than Secular and traditional users in several activities. These include Communication (4.2 times higher), Kids (2.9 times higher), Forums (2.6 times higher), YouTube (2.3 times higher) and E-Gov (1.4 times higher). The right-hand side of this "scissor effect" figure highlights the activities which are dominated by Secular and traditional users. These include Real-Estate (3.6 times higher), Adult (3.0 times higher), Dating (2.8 times higher), Rights Realization (2.5 times higher) and Social-networks (1.5 times higher). No substantial differences can be observed in the following activities: Search, Entertainment, Gambling, News, Sport, Services, Transportation, Finance, Public-service and Education.

Table 14 describes differences in content usage, broken down by the level of religiousness. As can be seen from the table, significant differences in mean visits can be identified in several categories. Secular and traditional users exhibited higher mean visits than religious users in the Adult (P<0.05), Social networks (P<0.001) and Travel (P<0.05) content usage categories. Religious users exhibited higher mean visits than secular and traditional users in the Communications (P<0.01) Kids (P<0.05) and Forums (P<0.05) content usage categories. The high usage volume in the latter category by religious users could be explained by the unique on-line behavior of the ultra-orthodox community (e.g. the need to "consume" content, access information and communicate using "Kosher" websites). Due to the special structure of this community and the restrictions it imposes on the use of the internet, special forums are often used as a substitute for social networks.

Discussion

This chapter presents an analysis of overt on-line user behavior. Unobtrusive data from various sources, parsed with reference to socio-demographic and locational attributes were used to analyze digital gaps in Israel. The main objective of this exercise was to provide a "proof of concept" for the study and analysis of the digital divide using digital traces. Although the temporal level and the scope of analysis was rather limited, encompassing only a tiny fraction of the trace data present in infinite digital space, the exercise was able to demonstrate the usefulness and application of this unobtrusive method and to produce other novel insights regarding the digital gap. The summarized information reported in Table 15, compares between the findings of our digital trace research (based on the SimilarWeb and Ifat panel analysis) and various findings from self-report studies. A close look at the table reveals a high degree of compatibility between the results of the digital trace exercise and findings reported in the digital divide literature.

As can be seen from the table, digital gaps in online behavior between female and male users was found to be substantial in both types of data sources in the following content usage categories: Information and Search; Entertainment; Finance; Dating (dominated by males) and Health (dominated by female).

Table 15: Content usage differences, parsed by socio-demographic attributes: Comparison of digital trace data (SimilarWeb and Ifat) with self-report findings

	(a) Content usage category	🟧 Higher usage ▨ Neutral 🟦 Lower usage			🟩 Yes 🟨 Partially 🟥 No
		(b) SWeb	(c) Ifat	(d) Self-report studies	(e) In-line with literature?
Gender **Female** (opposite color in columns b,c and d for reference group – male)	Information and search	🟦	🟦	🟦	🟩
	E-mail		▨		
	Health	🟧	🟧	🟧	🟩
	Government and rights realization	▨	▨		
	News	🟦	🟦		
	Sports		🟦		
	Work, career, research and education	▨	🟧		
	Entertainment (music, video and gaming etc.)	▨	🟦	🟦	🟩
	Communication tools, IM, chat and social networks	🟦	🟧	🟧	🟨
	Finance, commerce and business	🟦	🟦	🟦	🟩
	On-line shopping	🟦	🟧		
	Travel and tourism	▨	🟧		
	Transportation	🟦	🟧		
	Dating		🟦	🟦	🟩
	Gambling		🟦		
	Internet and telecom		🟦		
	Casual browsing				
Age **Young users** (opposite color in columns b,c and d for reference group – older users)	Information and search		🟦		
	E-mail		🟦	🟦	🟩
	Health		🟦	🟦	🟩
	Government and rights realization	🟧	🟦		
	News		🟦		
	Sports		▨		
	Work, career, research and education	🟧	🟧		
	Entertainment (music, video and gaming etc.)	🟧	▨	🟧	🟩
	Communication tools, IM, chat and social networks		🟦	🟧	🟥
	Finance, commerce and business	🟦	🟦		
	On-line shopping		🟦	🟦	🟩
	Travel and tourism	🟦	🟦		
	Transportation		▨		
	Dating		🟦		
	Gambling		🟧		
	Internet and telecom		🟧		
	Casual browsing		🟦		

Education Low level (opposite color in columns b,c and d for reference group – high level)	Information and search		▨		
	E-mail		🟧		
	Health		▨	🟦	🟢
	Government and rights realization		🟦	🟦	🟩
	News		🟦	🟦	🟩
	Sports		▨	░	░
	Work, career, research and education		🟦	🟦	🟩
	Entertainment (music, video and gaming etc.)		🟧	🟧	🟩
	Communication tools, IM, chat and social networks		🟧	🟧	🟩
	Finance, commerce and business		🟦	🟦	🟩
	On-line shopping		▨		
	Travel and tourism		🟦		
	Transportation		▨		
	Dating		▨		
	Gambling		🟧	🟧	🟩
	Internet and telecom		🟧		
	Casual browsing			🟧	
Income and socio-economic status Low level (opposite color in columns b,c and d for reference group – high level)	Information and search		▨	🟦	🟢
	E-mail		▨		
	Health		▨		
	Government and rights realization		🟧		
	News		🟦		
	Sports		🟦		
	Work, career, research and education		🟧	🟦	🟥
	Entertainment (music, video and gaming etc.)		🟧	🟧	🟩
	Communication tools, IM, chat and social networks		🟧	🟧	🟩
	Finance, commerce and business		🟦		
	On-line shopping		▨		
	Travel and tourism		🟦	🟦	🟩
	Transportation		🟦		
	Dating		🟧		
	Gambling		🟧		
	Internet and telecom		🟧		
	Casual browsing				

Substantial age-based differences in online behavior were identified, both in digital trace sources and self-report studies, in the following content usage categories: E-mail; Health; On-line shopping (dominated by older age cohorts) and Entertainment (dominated by younger age cohorts).

Substantial education-based differences in online behavior were identified in the following content usage categories: Government and rights realization; News; Work, career, research and education; Finance (dominated by users with higher levels of education) and Entertainment (music, video and gaming etc.); Communication tools, Instant messaging, chat and social networks and Gambling (dominated by users with lower levels of education).

Significant differences in online behavior between high-income and low-income users were identified in the following content usage categories: Entertainment (music, video and gaming etc.); Communication tools, instant messaging, chat and social networks (dominated by users with lower levels of income) and Travel and tourism (dominated by users with higher levels of income).

For several content usage categories, the results of the digital trace exercise contradict findings reported in the literature with regards to the direction of the gaps. For example, in the Communication tools, instant messaging, chat and social networks category, the digital trace analysis points out to higher usage by older age cohorts, while self-report studies show the opposite. In the Work and career content usage category (including on-line search for jobs), the digital trace analysis illustrates higher dominance by lower income users, whereas the literature reports the opposite.

The use of digital trace data in the study of the digital divide enabled us to investigate the prevalence of gaps in a wide range of on-line activities and content usage themes. Such examination is not possible in a single study based on surveys and interviews due to the limitations of self-report methods (the need to base the analysis on a very large sample to capture a wide range of on-line activities by the survey's subjects; Relaying on reported (stated) behavior rather on actual (revealed) behavior; and the ability of the respondents to provide an accurate weight or measure with regards to their on-line activities over time).

The findings of this exercise enable to infer policy conclusions with regards to the mitigation of digital gaps in several themes. These mostly involve on-line activities pertaining to every-day life such as **Health** (conducting on-line e-health activities such as making doctor appointments, requesting on-line prescriptions, searching for information on health-related issues); **Finance and real-estate** (e.g. information on pension and provident funds, conducting on-line banking transactions, investments etc.); **Rights realization** (e.g. accessing information on social security benefits, maternity rights and payments, minimum wage, disabilities etc.); **E-Government** (e.g. use of various e-gov services such as passport renewal, information on income taxes, paying local taxes such as water and property tax/arnona, land and property registration etc.) and **Gambling.** The results of this study reveal significant differences among various socio-demographic groups in the use of on-line activities. Relevant policy implications which can be inferred from these findings will be discussed in detail in Chapter 7.

Chapter 4 - Triangulation of Digital Trace Data: The Social Rights Realization Case Study

The objective of this chapter is to formulate a methodology for triangulating various digital trace data sources in order to deepen our understanding of the digital divide phenomenon and to construct more robust methodological tools, allowing evidence-based evaluation of actions. Triangulation is a commonly used approach, both in case studies and mixed methods research. In this approach, findings from one method are cross-validated by those in another with the aim of achieving greater validity in the research. Denzin (1978), who advocated a multi-source approach, defined triangulation as “the combination of methodologies in the study of the same phenomenon”.

In this study, we demonstrate the triangulation methodology by focusing on a rights realization case study in the context of the digital divide. Social rights, as stated by the UDHR[4], relate to various entitlements, such as the right for social security (Article 25 of the UDHR[4]), the right to work, the right to receive various work and employment benefits (e.g. employee rights as specified in Article 23 of the UDHR[4], favorable working conditions), the right to pursue adequate standard of living and the right for education (Barak-Erez and Gross, 2007). The UDHR determines that: “Everyone, as a member of society… is entitled to realization, through national effort and international co-operation”.

We choose to focus on rights realization and entitlements for two main reasons: First, rights realization is an important goal by itself, as declared by various stakeholders (Digital Israel 2017). Second, providing accessibility to rights realization information, especially to underprivileged and disadvantageous populations, might facilitate a better realization of rights and contribute to the reduction of social and economic gaps (Weiss-Gal and Gal, 2009; Benish and David, 2017).

This chapter is organized as follows. We first introduce the specific data and methodology used for the case study. Then, we present data stories which we find to be interesting and relevant to the digital divide issue. Finally, we discuss limitations and insights that were inferred from the study of rights realization obtained from digital trace data.

[4] http://www.un.org/en/universal-declaration-human-rights/

Social Rights Data

In this case study, we focus on two types of rights and entitlements realization: **employee rights** and **life events rights**. We analyze four employee rights: advanced training fund (advanced training), pension insurance, convalescence pay and minimum wage. We also analyze four life event rights: statutory maternity pay and leave (maternity), unemployment compensation (unemployment), disability payment (disabilities) and senior citizen benefit (elderly). The considerations for choosing these specific rights realizations included salience in rights non-realization (State Comptroller Report, 2017), as well as legal changes that took place during the examined period (e.g. minimum wage increase). A special emphasis was placed on the Kolzchut[5] and the National Insurance Institute of Israel (NII[6]) websites, as they constitute the major sources related to rights realization in Israel.

We use five digital trace data sources in our triangulation feasibility study: Buzzilla, digital traces of NII and Kolzchut websites extracted from the Ifat panel data subset, SimilarWeb tool, Google Trends and Google Analytics (GA). These data sources are described in detail on chapter 2. Specific data properties relating to the rights realization case study (e.g. extraction methods, metrics and time period covered) are summarized in Annex 5. A general description of data extraction methods pertaining to rights realization in the various data sources are presented below:

- **Buzzilla** - This tool enables to analyze **web-conversations** using user-defined strings. Linguistic Boolean **queries** containing relevant strings were developed for each of the nine analyzed entitlements. The query development process included manual fine-tuning iterations according to the relevance of the extracted conversations. A conversation item is typically composed of the following fields: title, section, date, social media channel, website and link. We particularly focused on qualitative text processing of two entitlements: maternity rights and elderly rights. The data was downloaded to excel files and additional fields were added: Indication for gender and indication for whom the information is requested or asked for (e.g. for personal use or for another person).

- **Ifat Panel Data Subset** - The Ifat data subset includes time-stamped desktop user entries to Kolzchut and NII websites from Israel, dating from 15/10/17 to 14/11/17. The various

[5] http://kolzchut.org.il
[6] https://www.btl.gov.il

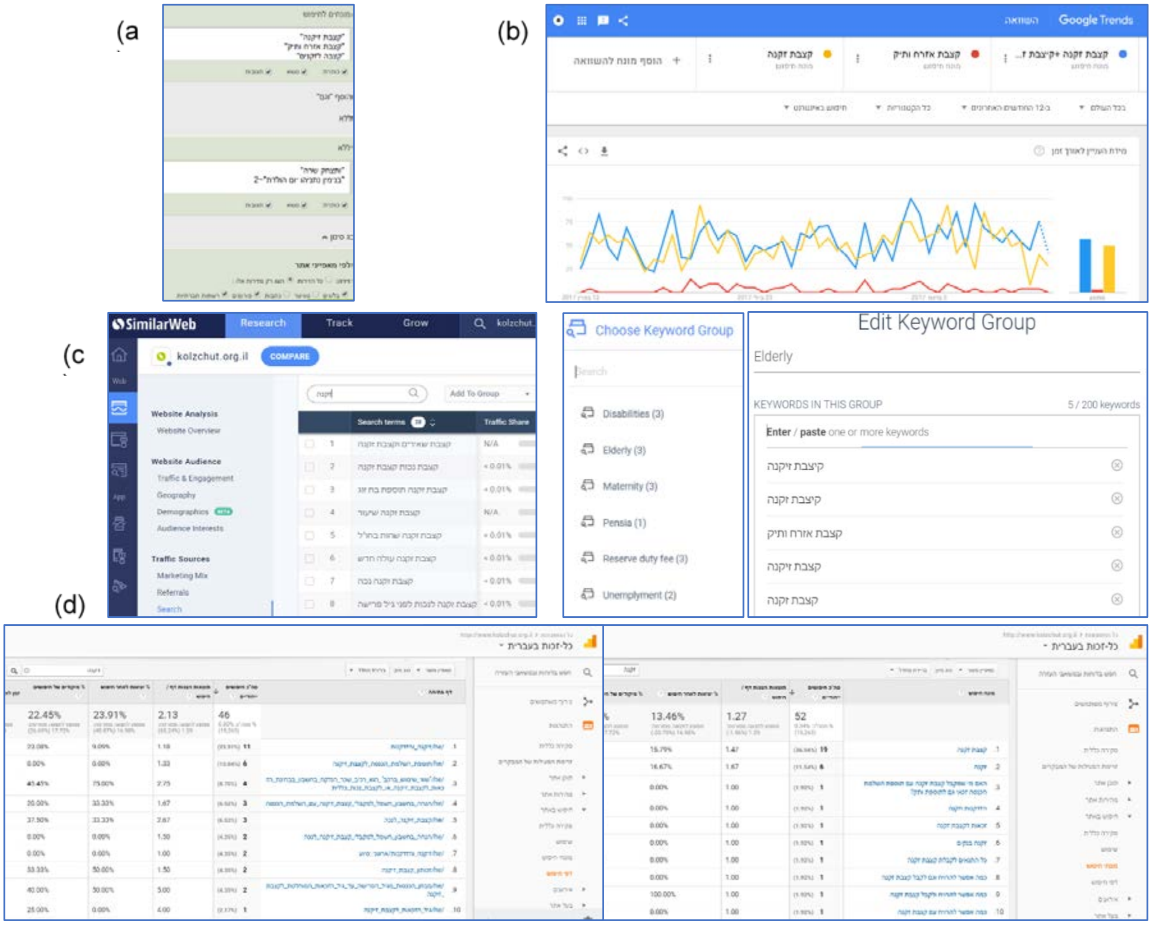
types of rights were identified and defined by **categorization of the landing page titles**. Processing page titles involved unicode conversions of full URL due to Hebrew language use.

- **SimilarWeb** – In this platform, two “off-the-shelf” tools were used: “Website Analysis” and “Search Keyword Analysis” (the latter is limited to desktop use only). The time range of the data extraction was set to October 15th, 2017 – November 15th, 2017 and the location was set to Israel.
  - **Website analysis** - Kolzchut and NII websites were both analyzed. The entitlements and rights were defined and differentiated according to the landing page title and the search terms associated with each right (Figure 12 c on left). For this purpose, daily data was downloaded (available on traffic and engagement report) and categorized. The categorization process was facilitated by excel tools. It is important to note that SimilarWeb provides a filter interaction within the search terms that lead to a specific website (available on search report in traffic sources section, limited to monthly data). This feature facilitates a categorization process as well.
  - **Search keywords analysis**- keywords and groups of keywords which identify each right are the subjects of the analysis (Figure 12 c on right). It is important to note that actions done by the users after obtaining the search results are unknown.

- **Google Trends** – A group of search terms was defined for each entitlement (Figure 12 b) using the “+” operator.

- **Google Analytics** – SNI was granted access to the GA of Kolzchut by the website administrator. The various rights were defined by the categorization of the landing pages titles (Figure 12 d on left). It is important to note that keyword searches leading to the website could be also used for analyzing the website (Figure 12 d on right). Demographic breakdown provided by the GA tool was used in order to segment reports by gender and age.

Figure 12: An example of a single social right application by various sources[7]



Methodology

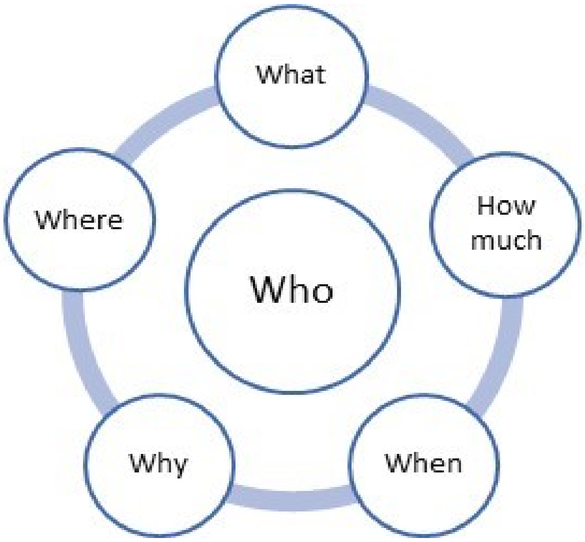
Figure 13 presents a set of guiding questions that were systematically used in the analysis of rights realization: What, how much, where, when, why and who. The six-question framework guided us in the study of each tool and data source. The goal was to associate features related to these questions to the context of the digital divide.

- **What** - the subject of the study, the unit of the research and the search terms used.
- **How much** – the scope of the activity in terms of visit and visitor volume.
- **Where** – the type of social activity (e.g. forums, blogs, social networks), the type of device used (desktop vs mobile) and the physical location of the activity (home vs. work).

[7] a: **Buzzilla** - elderly query definition; b: **Google Trends**- elderly search terms trends. The red colored term is in lower use than the yellow colored term. The blue colored term represents unification of the two other terms; c: **SimilarWeb** – filtering elderly search keyword on Website Analysis (on left), rights realization keywords groups on Keywords Analysis (middle), elderly group by keywords unification (on right); d: **Google Analytics** – filtering elderly landing pages (on left), filtering elderly search term (on right).

- **When**- the temporal dimension of the online activity (e.g. hours and days of search activity, web conversations etc.).
- **Why** – the reason or the motivation of a user for visiting a particular website or participating in a particular online activity.
- **Who** – the socio-demographic attributes of the on-line users (e.g. gender, age, Income level, geographical location, language etc.).

Figure 13: Six-questions social rights analysis framework for digital trace data sources



The "Who" question, which deals with the characteristics of the users, is the most relevant question to the study of the digital divide. When this question is used in tandem with the other questions: How much, where, when and why, we can make further steps towards investigating and studying gaps in on-line behavior between various populations.

Findings

In this section we present the main findings which are most relevant to the digital divide phenomena. The findings are presented in the format of seven data-driven stories demonstrating the rights realization case-study: the gender and age differences story, the mediators story, the "how-much" story, the "time-range" story, the naming story, the social media story and the "buzz" story. Data-driven stories are narratives that are based on data evidence, often portrayed by data visualization (Riche et al., 2018). All stories, except the "buzz" story, involve the use of triangulation methods. Annex 5 summarizes the social rights case study stories, as well as suggested policy guidelines for stakeholders and researchers. The tables in Annex 6 to Annex 11 summarize our technical use of trace data tools experience regarding our six-questions framework (presented in Figure 13).

Gender and Age Differences Story

The first data story presents evidence for the existence of gender and age differences in the context of rights realization. The metrics used are the number of visits and the number of unique visitors in the Kolzchut and NII websites, as reflected by several digital trace data sources: Ifat, Google Analytics and SimilarWeb. As can be seen from Figure 14, slightly lower online activity (as exemplified in the number of web visits and in the number of unique visitors) exists among female users.

Figure 14: Digital trace data exemplifying gender differences in online activity in the context of rights realization[8]

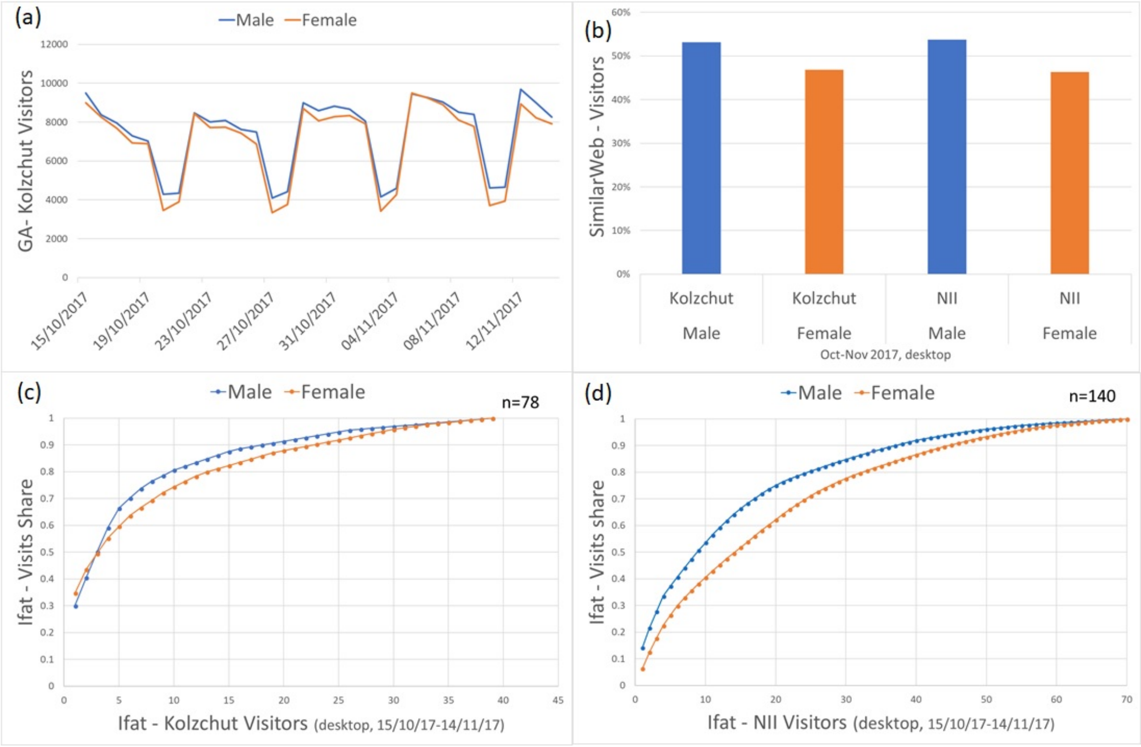
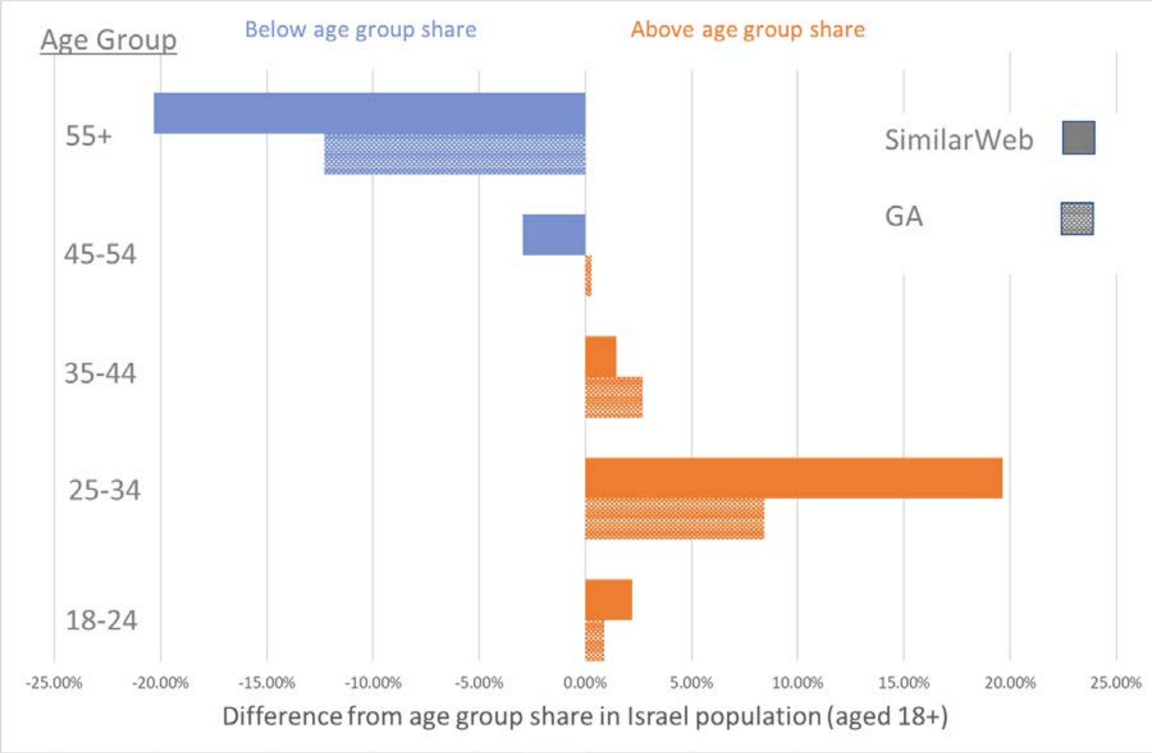


Figure 15 presents differences in the on-line activity share of various age groups (the metric used is the number of unique visitors in the Kolzchut website) in the Google

[8] a: **Google Analytics-** unique users of Kolzchut website (using all devices). b: **SimilarWeb -** unique desktop users of Kolzchut and NII websites. c: **Ifat data-** a cumulative histogram of desktop entries to Kolzchut website. d: **Ifat data-** a cumulative histogram of desktop entries to NII website.

Analytics and SimilarWeb platforms. The differences were normalized by accounting for the representative share of these age groups in the Israeli population (CBS, 2016).

Figure 15: Digital trace data exemplifying age differences in online activity in the context of rights realization[9]



As can be seen from the graph, both the Google Analytics and SimilarWeb platforms exhibit similar trendline with respect to the behavior of the users, whereas the 25-34 age group is the most active in terms of rights realization (e.g. unemployment and work-related rights, maternity rights etc.). This finding is consistent with the results presented in chapter 3, showing higher use by the 25-35 age group. As can be observed from the data, rights realization decreases with age. ANOVA tests for differences in mean visits in the Kolzchut website (obtained from GA) show a highly significant main effect for age (*p*<.05). Post-hoc analysis for visit means using the Bonferroni Correction show that significant differences (*p*<.05) exist between all age groups, excluding the 45-54 and the 55-64 age groups. In terms of Kolzchut visitors, a highly significant main effect for age (*p*<.05) was also observed. Post-hoc analysis for visitor means using the Bonferroni Correction show that significant differences (*p*<.05) exist between all age group pairs.

[9] Time period of online activity: Oct-Nov 2017. Population distribution by age data taken from Central Bureau of Statistics website.

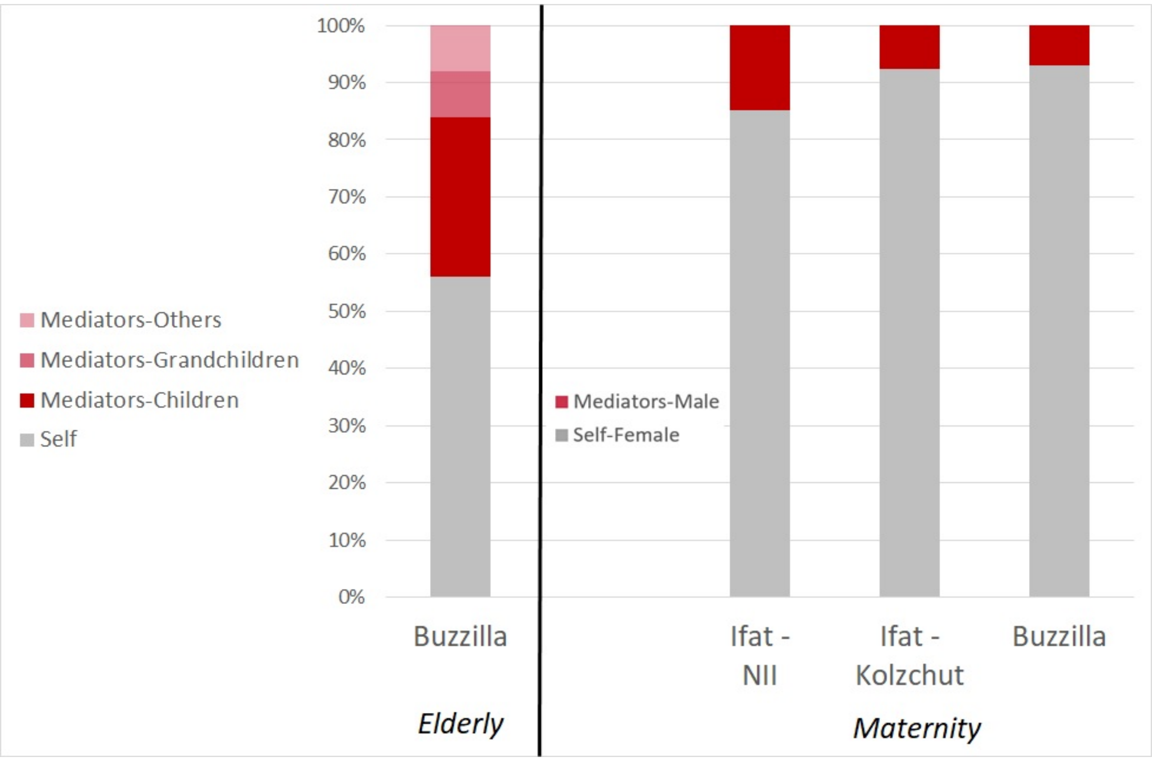
The “Mediators Story”

A close look into rights realization conversations in the Buzzilla platform reveals two types of “players”: a person who asks for information for himself and a person who asks or requests information for others. The latter “player” is referred in the literature as a “mediator” (Benish and David, 2017). The “mediators story” is highly relevant to digital divide research, as it indicates cases where people do not use the internet themselves and require help from others.

The Hebrew language enables one to identify the subject’s identity (in terms of “self” vs. “mediator”, e.g. whether individuals ask the question for themselves or whether they mediate for others). The gender of the subject who asks the question can be identified by verbs declension. For performing this task, manual screening is required. During the investigation of the “Mediators Story”, we examined all conversations regarding elderly and maternity rights performed between 15/10/17 and 14/11/17. Among 25 conversations that were found relevant to the elderly rights realization topic, 44% involved the work of mediators (for example children on behalf of their parents, grandparents or others, see Figure 16 on left).

The analysis of the maternity rights realization topic reveals an unmediated feminine discourse (see Figure 16 on right). Apparently, this maternity discourse is led by young women who do not require any mediation. Male subjects were found to be rarely involved in this type of discourse (e.g. in case their spouses do not speak Hebrew). The Ifat Panel data strengths the above findings and reveals strong indication for feminine discourse as well (Figure 16 on right). Google Analytics data for the same time period, focusing on maternity rights realization search terms, yields results showing that 100% of the search activity (for known gender) was conducted by females.

Figure 16: Digital trace data exemplifying mediators' role in the realization of rights[10]



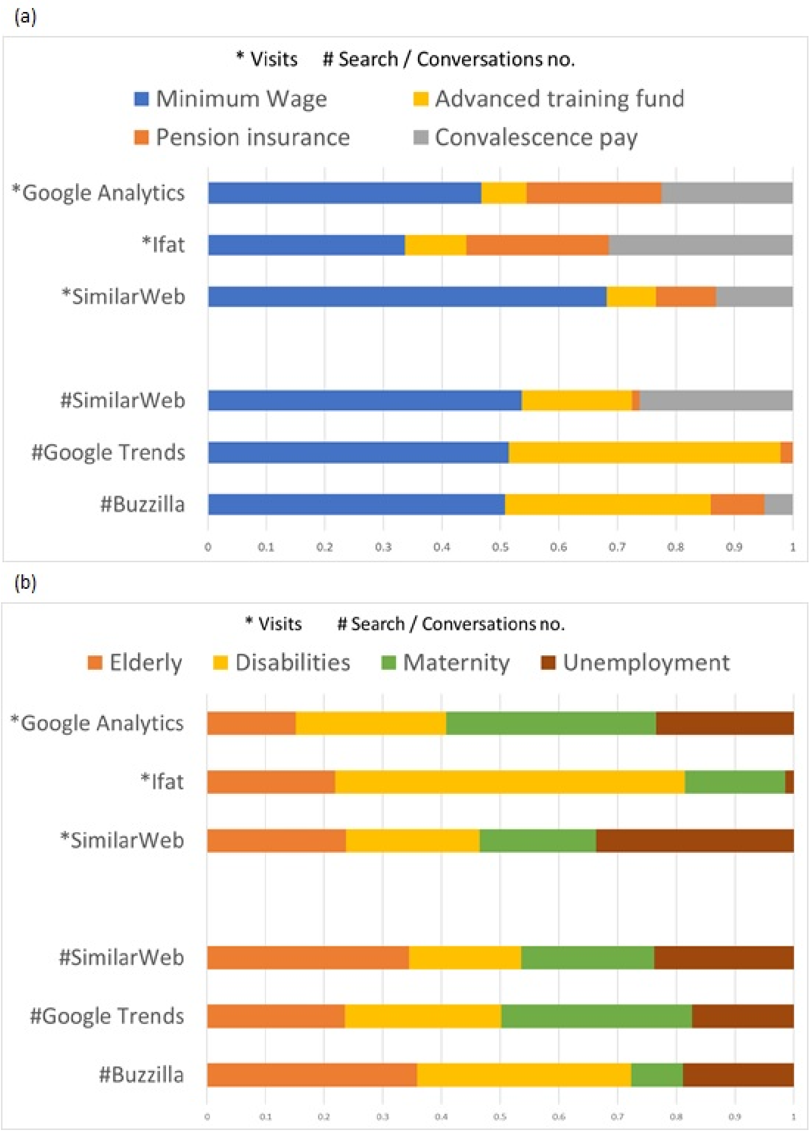
The "How-much Story"

In this data story, we analyzed the differences between five various data sources with respect to two types of rights realization categories (employee and life event rights), covering the same time-period. The scope of activity (measured in the volume of conversations, number of searches or visits, see Annex 7) in each rights realization category was compared across the five data sources.

Figure 17 presents the distribution of employee rights and life-event rights realization volume in five data sources. As can be seen from the figure, the minimum wage right dominates the employee rights realization category, whereas in the life-event category no clear dominant right could be observed.

[10] On left – Text analysis of conversations retrieved from Buzzilla. On right – Ifat's data analysis showing indication for significant female discourse as well.

Figure 17: Distribution of employee rights and life-event rights volume in five[11] data sources



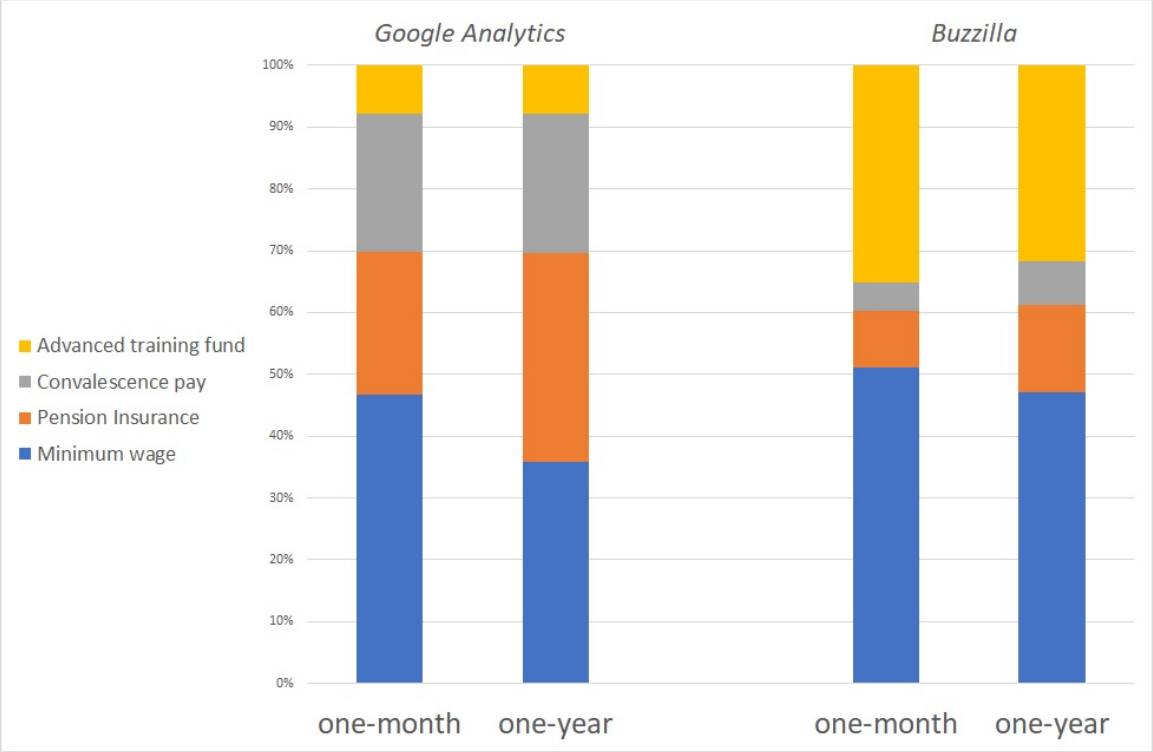
The “Time Range” Story

Figure 18 presents comparison of volume proportions in two time periods: a one-month period (15/10/17-15/11/17) and a one-year period (the year 2017). This exercise was conducted both in the Buzzilla and in the Google Analytics tools. The result of this examination shows a similar trend in the distribution of the rights over the two time periods. Relatively higher proportion in the minimum wage right could be observed in the one-month

[11] Two metrics in SimilarWeb.

period. This could be explained by legal changes that took place during this one-month period (a decision to increase the minimum wage).

Figure 18: Rights' volume - one-month vs. one-year[12]



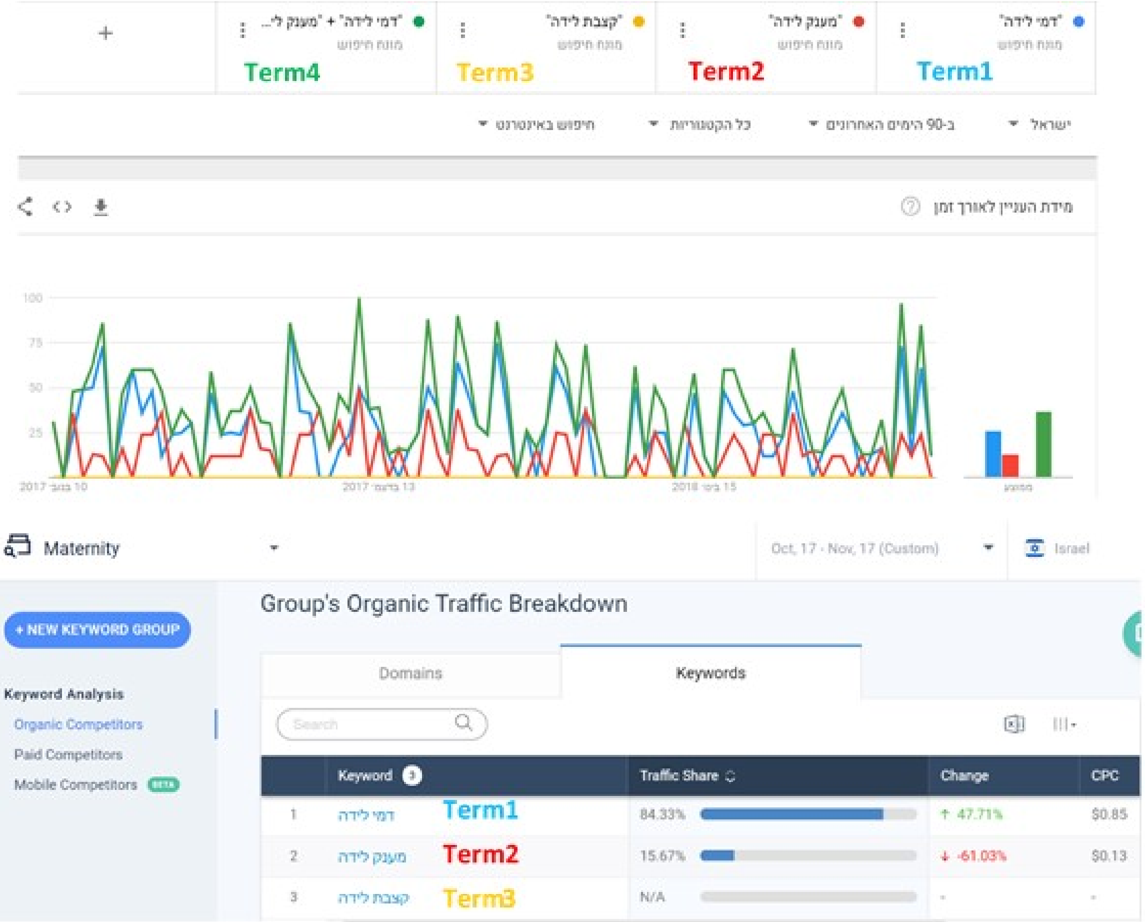
The "Naming Story"

The naming stage is defined as the ability to translate and accurately name a specific benefit (Felstiner et al., 1980). Naming an entitlement or a right is an important step required for its realization. Knowing the terms that are in use when searching for information on a specific benefit might help in raising its realization potential. Examining digital trace data concerning entitlements can facilitate an understanding of the naming stage process. This could be done using a variety of tools. The Google Trends tool and the Search Keywords Analysis by SimilarWeb scan **all** google searches, while the Google Analytics and the Website Analysis of SimilarWeb enable to analyze search terms that lead to specific websites.

[12] Volume metrics: Buzzilla - conversations no.; GA - Visits no. one-month: 15/10/17-14/11/17; one- year: 2017

The top section of Figure 19, presenting a Google Trends screenshot, shows two popular terms that are in use for Maternity benefit (colored blue and red). A third optional term (colored orange) is not in use. Combining keywords (e.g.: **Term4**=**Term1** or **Term2**) is an important feature which may further contribute to the examination of the naming process. The bottom section of Figure 19 shows the use of the same terms in the SimilarWeb tool.

Figure 19: Maternity benefit naming[13]



Group's Organic Traffic Breakdown

	Keyword 3		Traffic Share	Change	CPC
1	דמי לידה	Term1	84.33%	↑ 47.71%	$0.85
2	מענק לידה	Term2	15.67%	↓ -61.03%	$0.13
3	קצבת לידה	Term3	N/A	-	-

The Social Media Story

The five social media channels covered by Buzzilla are: social networks, forums, articles, Twitter and blogs. We analyzed these conversation channels for both employee rights and life-event rights. We compared the number of data conversations from the different channels over a one-month period. Figure 20 shows the distribution of conversations by

[13] On top - **Google Trends –** Search Terms**.** <**Term4**> = <**Term1**> or <**Term2**>. <**Term3**> is in very low use. On Bottom – **SimilarWeb** Search keywords. Retrieved on 15/1/18

social media channels for two employee rights (on top) and for two life event rights (on bottom). As can be observed, social networks (green color) constitute the most dominant social media channel for rights realization, while Twitter (orange color) and Blogs (light purple color) are the least dominant. These findings were found to be in line with the Bezeq Report (Bezeq, 2017). Further analysis shows that Twitter popularity in Israel is relatively low (In Feb-2018 Twitter's rank in Israel 31 vs. 13 in global ranking, https://www.alexa.com/topsites/countries/IL).

Figure 20: Comparison of four different rights realization conversations on Buzzilla social media channels[14]

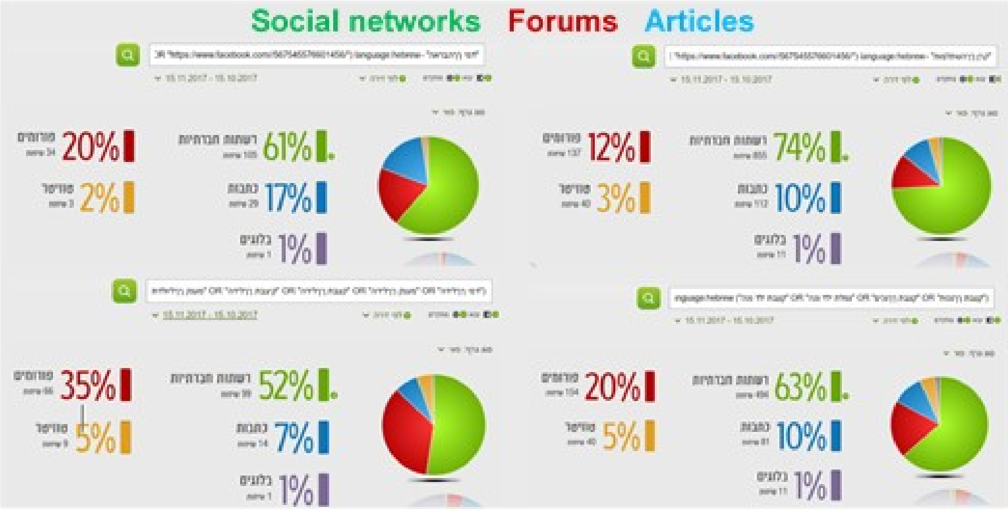
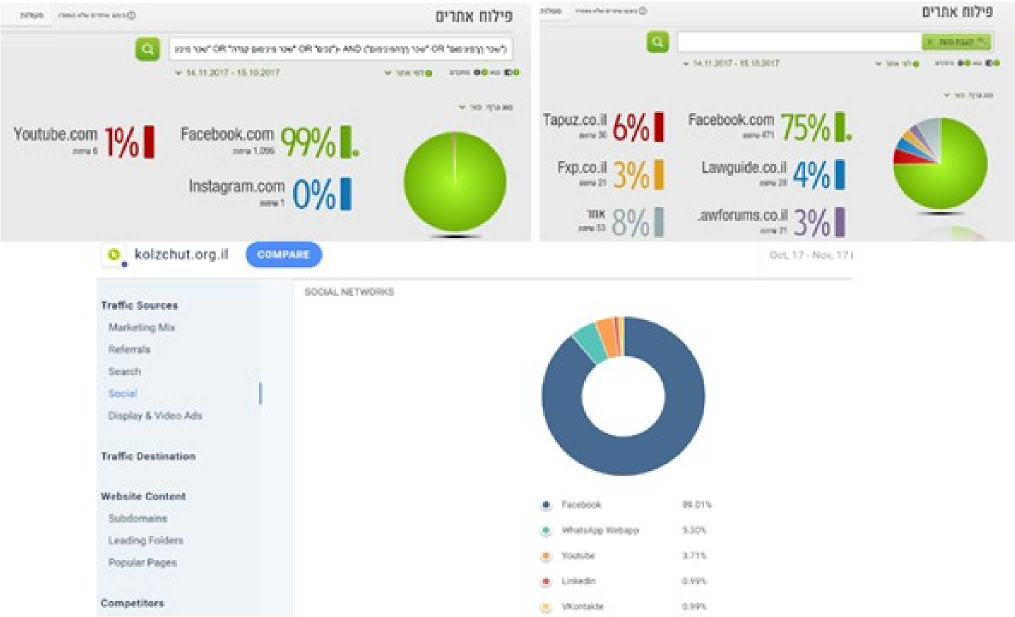


Figure 21 shows the distribution of conversations by websites. The top section, presenting Buzzilla screenshots, shows that Facebook leads in conversations in the topic of Minimum Wage rights (on left) and in the topic of disability rights (on right). The bottom section, presenting a SimilarWeb screenshot, demonstrates a relatively high use of Facebook as a platform or website leading to (referral) the Kolzchut website as well.

[14] Buzzilla screen shots. Retrieved on 15/1/18.

Figure 21: Minimum wage right volume by websites use distribution[15]

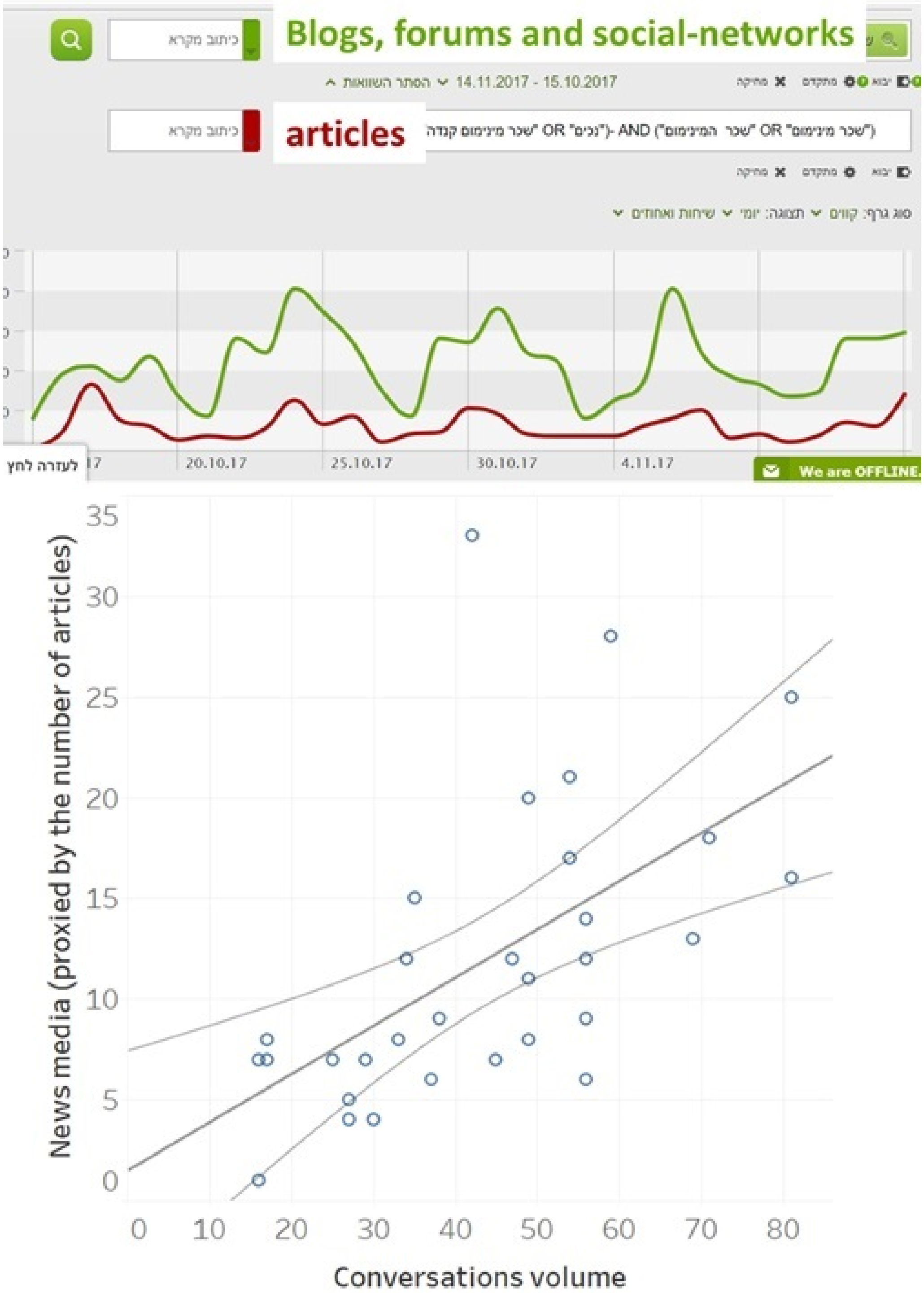


The “Buzz” Story

What is the timing of rights realization searches and conversations? What is the trigger? We expect news media to instigate public involvement, thus raising conversation volume (King et al., 2017). We use the Buzzilla tool to study correlation between news media volume (proxied by the number of articles) and conversation volume (in social networks, blogs and forums) regarding Minimum wage in Israel. An analysis of the Pearson correlation (see Figure 22 on bottom) indicates that a strong positive association exists between minimum wage articles volume and conversation volume on the subject ($r = .59$, $p = < .001$, $n = 31$). It is interesting to see that the peaks in article volume and conversation volume (Figure 22 on top) are connected to the timing of policy change (e.g. raising of minimum wage). These results stand in line with King et al. (2017), and demonstrate that in the rights realization domain, news media coverage prompt public interest, active involvement and contribute to public discourse.

[15] On top - Buzzilla screen shots – Websites distribution of conversations volume on Minimum wage (top left) and Disabilities (on right). On bottom - SimilarWeb screen shot assessing Facebook to be the dominant social referral website to Kolzchut. Retrieved on 15/1/18.

Figure 22: Minimum wage conversations volume during 15/10/17-14/11/17– articles vs. blogs, forums and social-networks[16]



("שכר מינימום" OR "שכר המינימום") AND -("נכים" OR "שכר מינימום קנדה

[16] On top - adding time dimension might shed light on association with policy changes. (Buzzilla screenshot, retrieved on 10/3/18). On bottom - A strong positive significant correlation between news media and public conversations was found.

To summarize, we presented seven data stories that were triggered by our rights realization case study. We showed evidence for the existence of gender and age differences in the realization of rights. We found evidence for web-mediators presence among elderly people. Slight indications of similar volume proportions across sources were found. Indication for similar rights volume proportions in different time periods were found as well. We demonstrated the process involving entitlement naming by keywords analysis. In the social media story, we showed that individuals tend to discuss rights realization on social networks, particularly on Facebook. Finally, in the "Buzz" Story, we presented evidence, which stands in line with previous studies, showing that the news media constitutes "a trigger" for instigating public conversation. All stories (except the "Buzz" story) demonstrate the importance of the triangulation method in cross-validating trace data sources.

Chapter 5 - Visualization of the Digital Divide Using Trace Data

Visualizing the digital divide is highly important. The processes of understanding information and decision making are known to be affected by the way information is presented (Dickson et al. 1977; Elting et al. 1999). Digital divide visualization is important for raising stakeholder's interest, understanding and trust in the data (Cherchye et al. 2007). However, the role of digital divide visualization is much broader than merely providing a policy-analysis tool for stakeholders. In recent years, the process of improving transparency in the economic and social apparatus of the government is perceived to be vital for democracy. Public communication of the kind that enables transparency of data and citizens' involvement, has become a highly important goal (Cukier 2011). A key element in Big Data analytics is the presentation of the findings of the analysis in a user–friendly format (Pulse, 2016). Previous research about digital divide visualization was related to survey-based data (Albo et al., 2016; Albo et al., 2017). In this chapter we discuss the issue of digital divide visualization by using digital trace data.

First, we review data characteristics and their visualization design implications. Then, we describe relevant visualizations present in the tools used in the framework of this research. Finally, we discuss insights and lessons learnt from this research concerning digital trace-data visualization in the context of digital divide, as well as future work.

Trace Data in the Digital divide Context

Many aspects of visualization design are driven by the kind of data to be visualized. Digital trace data are human and machine "footprints" of digital activity. Data items, which are individual discrete entities, for example: likes, posts, blogs, search activity in search engines, transactions of e-purchases etc. Trace data might be organized in various levels (e.g. user-level, session-level or "hit"-level) and can be grouped by demographics (e.g. age, gender, joint age and gender groups like "young males"), or by affinity categories and market segments (e.g. "News Junkies", "Shoppers", "TV Lovers" etc.). The attributes of the items, which are specific item properties, might be either numerical (e.g. number of visitors in a specific website in a given time-period), or textual (e.g. conversations' text, search keywords or landing page titles).

Trace data in the digital divide context is characterized by high volume of multi-dimensional time-oriented data. Another aspect to consider is the challenge of organization and integration of multiple sources of data (Kourtit and Nijkamp, 2018; Pulse, 2016). In the following section, we explain each term in detail.

Multidimensionality

This property concerns the number of variables in the data. In principle, it makes a difference if the data is represented by a single value or by multiple values. The multidimensionality property of digital trace data in the context of the digital divide is a consequence of two factors:

1. Multiple metrics

Trace data reflect internet use by several measurements. For example, the scope or intensity of using a particular website can be measured by metrics such as the number of unique users, the number of sessions, the share of new sessions, the duration of the session etc. A conversation in the internet might be characterized by the media channel used (e.g. social networks, forums, blogs etc.), the specific website in use (e.g. Facebook or Twitter) or by the frequency of keywords in a specific category (e.g. number of positive versus negative keywords in sentiment analysis). Each metric presentation might support digital divide exploration.

2. Multiple concept dimensions

Information and Communications Technologies (ICT) readiness, which is a complex realty, is the phenomena studied to point out digital divides between or within populations. Thus, ICT measurements reflect multiple dimensions (e.g. internet infrastructure, use and impact). Specifically, internet use is a multidimensional concept (Blank, 2014). Internet activity can be sorted by content usage categories, such as e-Shopping, e-Health, e-Gov, email, online banking, gambling etc. Categorization of trace data to content usage categories might involve manual or automated lexical analyses processes (Schober et al., 2016). Relevant questions regarding multiple dimension tasks in the context of the digital divide may include the following: Do digital divides exist between entities (as between gender or between age groups); Is there a difference between the type of online content consumed? Do differences exist between entities with respect to single or multiple category use? (e.g. single category: differences between age groups consuming e-health services; multiple category: differences between age groups consuming e-health, e-gov and finance services).

Time-Oriented Data

Trace data, which is basically the reflection of users' behavior in the internet, is a time-stamped data. Internet activities as sessions or conversations are usually documented by date, hour, minutes and seconds.

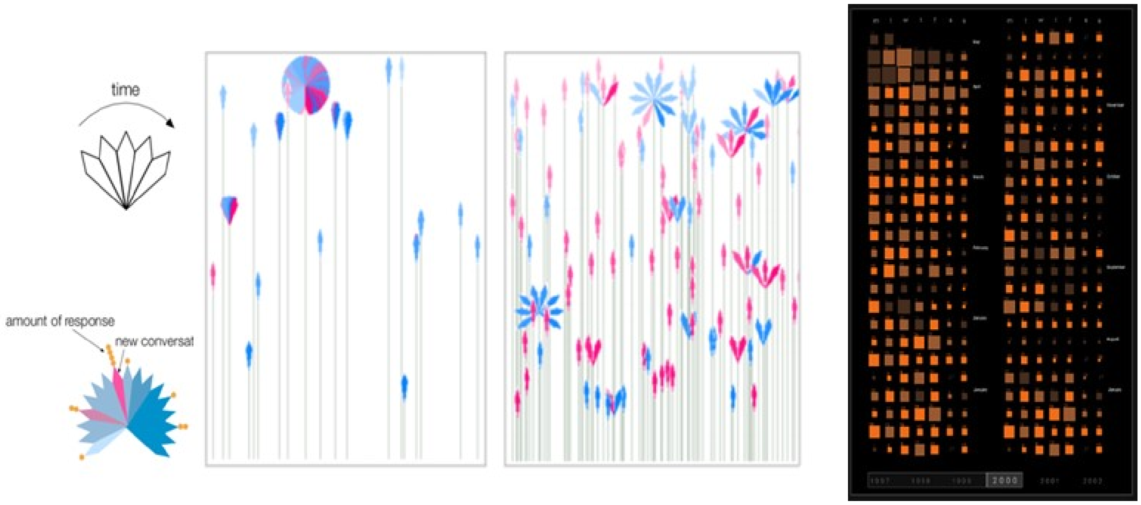
In addition, ICT measurement and benchmarking processes are usually continuous procedures. Supporting time-oriented data and tasks are one of the core challenges of visualization in the digital divide domain. Time-oriented questions related to digital divide exploration might be: What is the direction and scope of the divide – does it diverge or coverage? What is the rate of change? (Sciadas, 2004). These questions relate to data trends which are likely to be aggregated, in linear arrangement of the time dimension (i.e. from the past to the present). However, exploring temporal patterns in internet use might be interesting as well (e.g. exploring e-shopping patterns during parts of the day, week day or month). In this case, temporal cyclic arrangement can be considered (i.e. time is organized in a set of recurring time values). The combination of periodic and linear trends (e.g. seasonal/monthly e-shopping over a few years period) is denoted by the term "serial periodic data" (Aigner et al., 2014).

Aigner et al. (2014) introduce a survey[17] of 155 visualization techniques for time-oriented data, 70 of them are relevant for multidimensional data. Figure 23 shows two examples of multivariate time-oriented data visualizations, with both linear and cyclic time arrangement. These examples, which focus on on-line data, do not deal directly with internet use. However, they might provide ideas for innovative methods for multivariate time-oriented trace data visualization and inspire visualization designers in the internet use domain. PeopleGarden (Xiong and Donath, 1999) visualizes data on users and messages posted on an online interaction environment (see Figure 23 on left). Users are represented by flowers whose petals represent individual messages posted by a user. The platform integrates information on the time of posting, amount of response, and whether a post starts a new conversation. The garden represents the whole environment. The height of a flower represents how long a user has been in the interactive environment.

[17] http://timeviz.net/

PostHistory (Viégas et al., 2004) is a user-centric system that was developed with the goal of visually uncovering different patterns of e-mail activity and the role of time in these patterns. Its calendar panel shows e-mail daily activity, where the volume of the activity (e.g. the number of emails sent or received) is mapped to a box size, and its average directedness (i.e. whether a mail was received via TO, CC or BCC) is highlighted by the brightness level (see Figure 23 on right). In short, PostHistory presents a personal portrait of an individual through the context of their interactions, providing an accessible way of looking at high-level patterns of email activity over time.

Figure 23: Examples of multivariate time-oriented data visualizations with linear and cyclic time arrangement[18]



Source: PeopleGarden: Xiong and Donath, 1999 (on left); PostHistory: Viégas et al., 2004 (on right)

Multiple-Source Data

Trace data research often deals with data derived from multiple sources (Hampton, 2017; UN Global Pulse, 2015). Integration of multiple sources data might be challenging. Visualization techniques might be powerful in transforming digital interaction traces into interpretable forms by associating and synchronizing the different sources (Laflaquière, 2009). However, the role of digital trace data visualization is crucial when it is in fact the major, and sometimes the only method that enables integration of multiple trace data sources.

[18] On the left – PeopleGarden. A group with a dominating voice vs. a more democratic group. On the right – the calendar panel of PostHistory.

A possible design solution for multiple-source data would be to split up the display into multiple views. Few (2007) defines a “faceted analytical display” as “a set of interactive charts (primarily graphs and tables) that simultaneously reside on a single screen”. He distinguishes between dashboards which are used for “monitoring what’s going on” and displays that “combine several charts on a screen for the purpose of analysis”. Tableau[19] software provides visualization solutions for integrating several sources, as well as built-in API connections to trace data sources (e.g. Google Analytics).

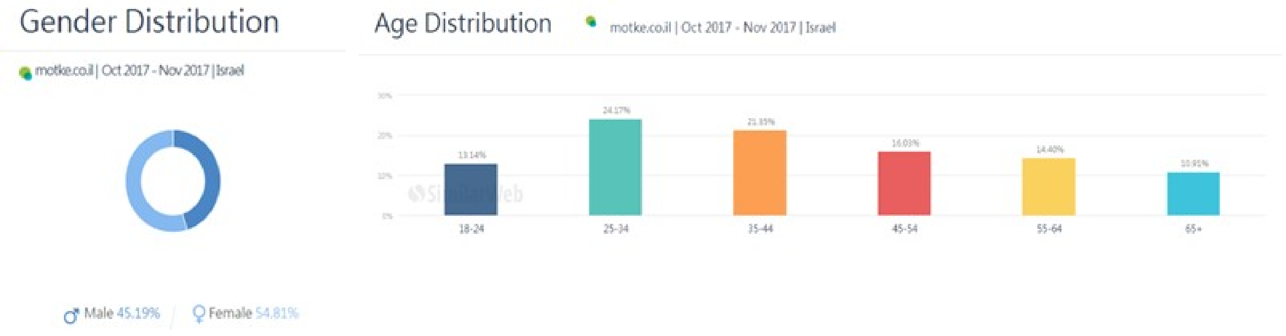
Visualization Tools Used in the Current Research

As the study of the digital divide sets on comparing populations on a demographic base, we focus on the tools that provide demographic reports. SimilarWeb and GA are two trace data tools used in the framework of this research which provide “off-the-shelf” visualizations.

SimilarWeb

SimilarWeb provides reports on the number of unique users with the ability to segment them by gender and age for at least a two-month period. Figure 24 shows SimilarWeb interactive chart reports for gender and age distributions for a specific website designed for senior citizens (motke.co.il). Gender share is represented by a donut chart, while age distribution is represented by a colored bar chart.

Figure 24: SimilarWeb visualization of gender and age distributions of a website intended for older users[20]



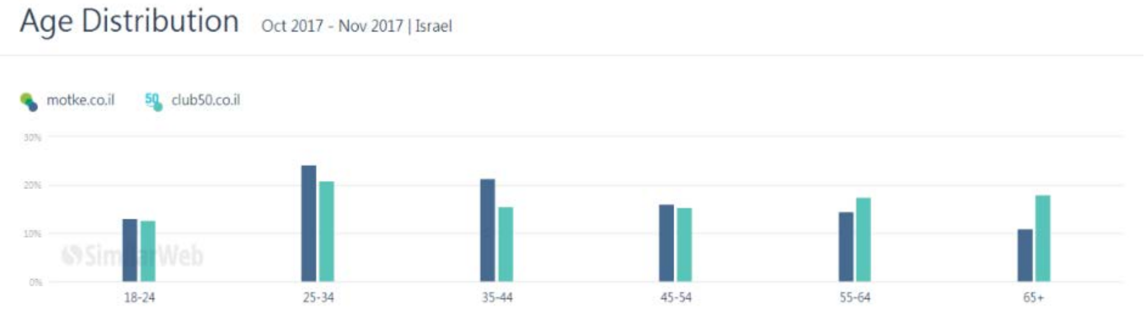
Comparison of the demographic distribution between two or more websites is also possible in the report described above.

[19] https://www.tableau.com/

[20] On left - Higher user distribution among female subjects can be observed. On the right - comparisons of age distribution for six age cohorts. Retrieved on 14/4/2018.

Figure 25 shows a bar chart illustrating a comparison of age distributions between two websites. Relative inferences about the scope of the on-line activity in these two websites could be made (e.g. the use of the club50 website among the 55+ age group is higher than the use of the Motke Website – see Figure 24).

Figure 25: SimilarWeb visualization of age distribution of two websites designed for senior citizens/older users[21]



Google Analytics (GA)

Several reports concerning the demographic characteristics of on-line users are provided by GA. Figure 26 shows a demographic overview report of GA, regarding Kolzchut use during the 15/10/17-14/11/17 time period. As visualized on SimilarWeb, Gender distribution on GA is also presented by a pie chart, while age distribution is represented by a bar chart.

GA provides the possibility to display interactions which significantly enrich the exploration platform. Users can select items to present (by the segmentation option), attributes (i.e. metrics) as well as time granularity and time range. In some cases, the user can select the chart types by which data is presented. In the following paragraphs, the types interactions are presented.

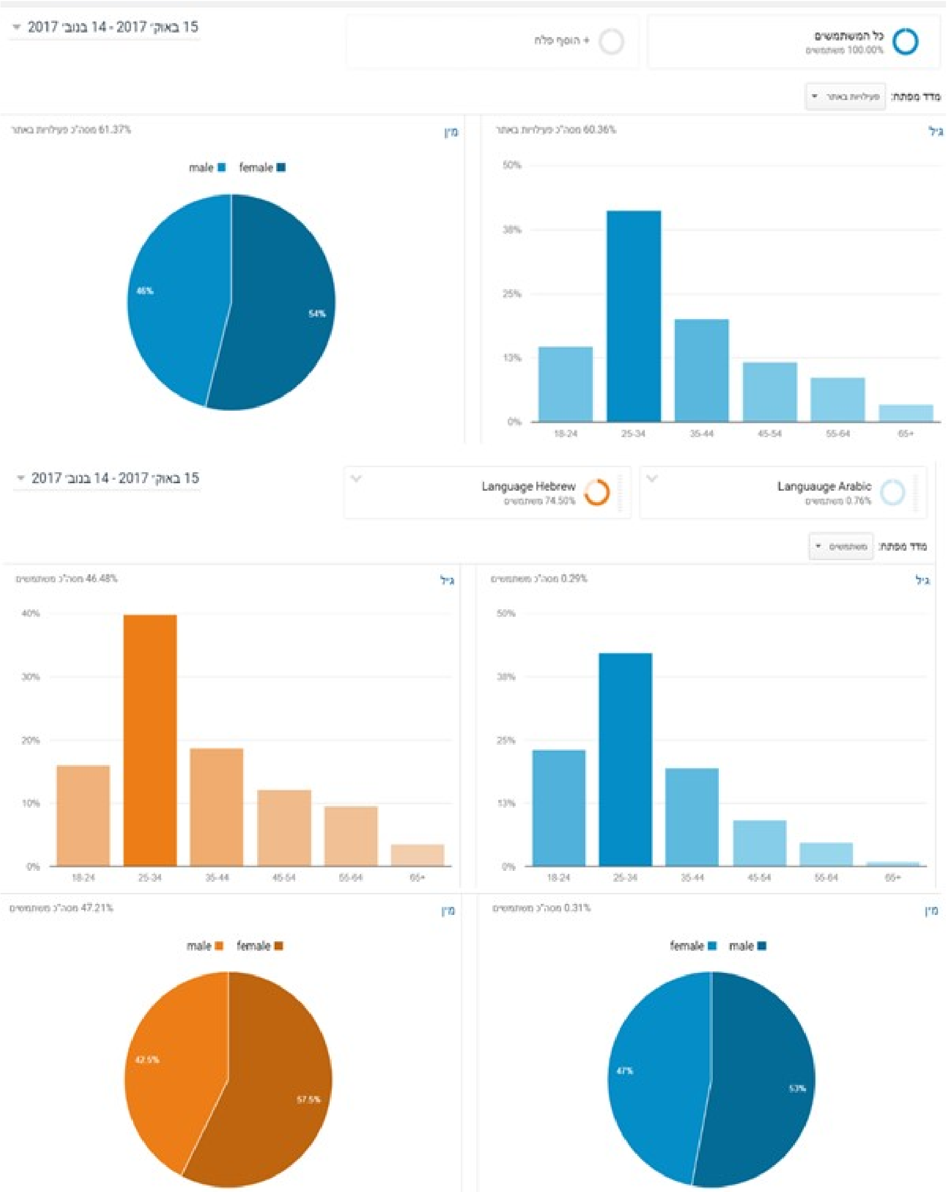
A segment is a subset of Analytics data[22]. It is composed of one or more filters. These filters isolate subsets of users, sessions and hits. In the digital divide context, we focus on subsets of users. The user filters could be identified by demographic attributes (e.g. gender, age, language and location) and by a combination of these attributes. For example, out of the entire set of users, the Hebrew speakers segment could be compared

[21] Higher use of club50 than motke among people aged 55-64 and 65+ can be seen, demonstrating relative sites use comparison. Retrieved on 14/4/2018.

[22] https://support.google.com/analytics/answer/3123951?hl=en

to the Arabic speakers segment (see Figure 26 at the bottom). Another segment might be users belonging to a specific gender, age group or a combination of these two demographic attributes (e.g. female aged 45-54). It is possible to create up to 100 segments per user.

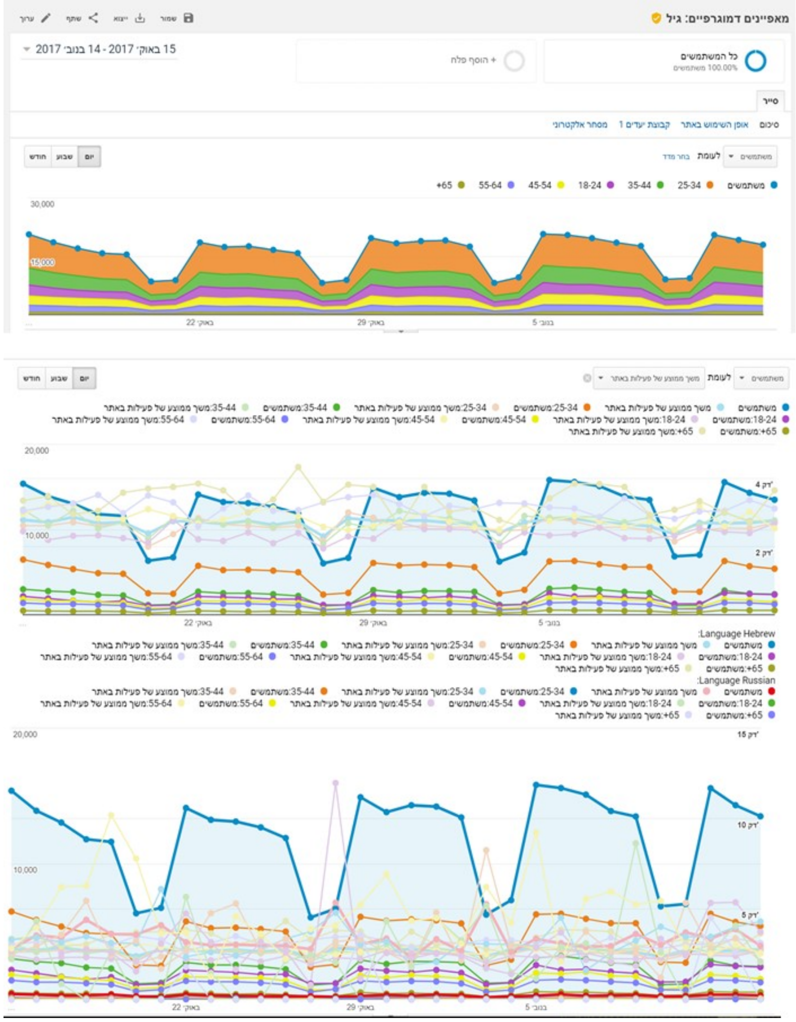
Figure 26: Google Analytics visualization of gender and age distributions of Kolzchut users between 15/10/17 – 15/11/17[23]



[23] On top – gender and age distribution of all Kolzchut users. Metric: Activities. On bottom - gender and age distribution of Arabic language users vs. Hebrew language users. Metric: unique users. Retrieved on 10/5/18.

GA also provides control over the presented metrics. A user can manipulate and change the metric used in the report (e.g. number of sessions, session duration etc.) and in some reports (e.g. age and gender reports), two metrics can be compared at one view. Time granularity can be changed to be either on daily, weekly or monthly basis. Figure 27 shows the demographic age report of GA for the Kolzchut website use during the 15/10/17-15/11/17 time period. Age distribution is presented by an area chart (stacked) and differentiated by color (see Figure 27 on top). The user can add a metric to create a comparison (Figure 27 on middle). Metrics comparison is shown by a line chart and is differentiated by color. The user can add comparison segments (Figure 27 on bottom).

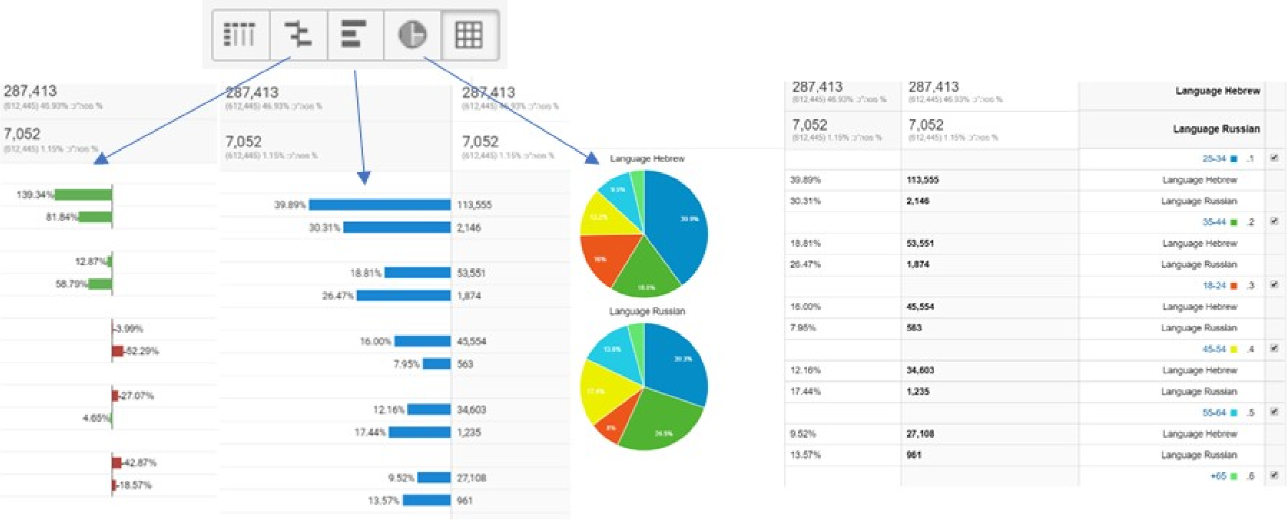
Figure 27: Google Analytics visualization of age distribution in Kolzchut website[24]



[24] GA age distribution report. On top –Metric: Activities. Population represented: All users. On middle - Metric: unique users vs. Average session duration. Population represented: Hebrew language users. On bottom - Metric: unique users vs. average session duration. Population represented: Hebrew language vs. Russian language users. Retrieved on 10/5/18.

Figure 28 shows a detailed tabular report which is presented below the age report. The user can select a pie chart, a bar chart or a diverging bar chart to be displayed along the table values.

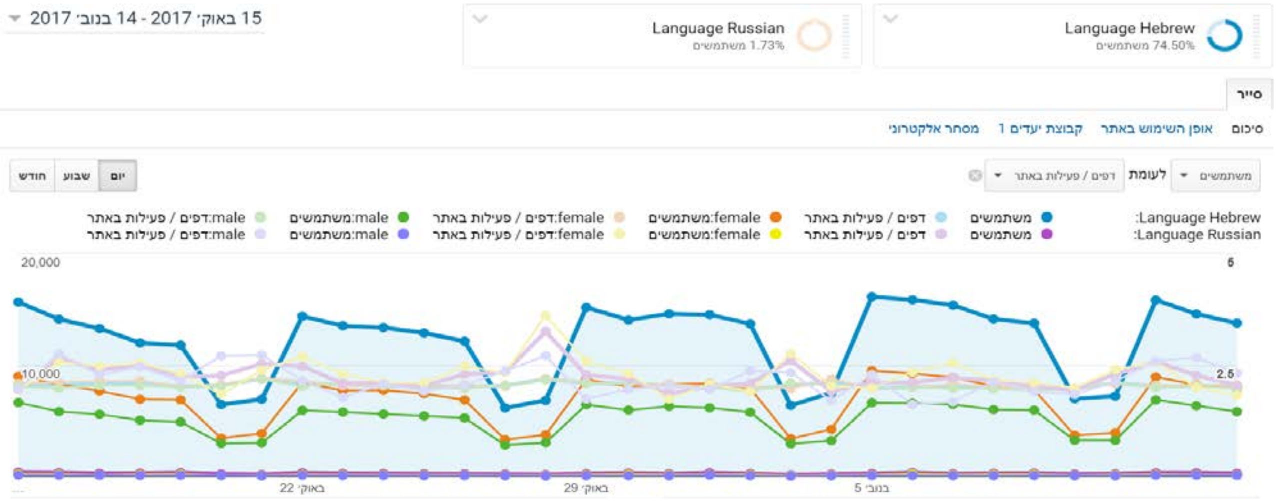
Figure 28: Google Analytics detailed age report for the Kolzchut website[25]



287,413	287,413	Language Hebrew	
7,052	7,052	Language Russian	
		25-34 🟦 .1	☑
39.89%	113,555	Language Hebrew	
30.31%	2,146	Language Russian	
		35-44 🟩 .2	☑
18.81%	53,551	Language Hebrew	
26.47%	1,874	Language Russian	
		18-24 🟥 .3	☑
16.00%	45,554	Language Hebrew	
7.95%	563	Language Russian	
		45-54 🟨 .4	☑
12.16%	34,603	Language Hebrew	
17.44%	1,235	Language Russian	
		55-64 🟦 .5	☑
9.52%	27,108	Language Hebrew	
13.57%	961	Language Russian	
		+65 🟩 .6	☑

Figure 29 shows the demographic gender report of GA on Kolzchut use during the 15/10/17-14/11/17 time period. This report features are similar to the age report described in Figure 27.

Figure 29: Google Analytics gender report for the Kolzchut website[26]

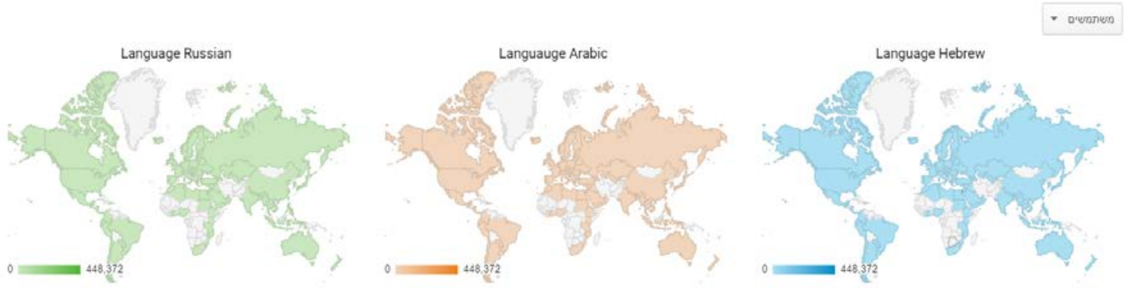


[25] GA detailed age distribution report. Metric: unique users. Population represented (segments): Hebrew language vs. Russian language users. Retrieved on 10/5/18.

[26] GA gender distribution report. Metric: unique users vs. Pages / Sessions. Population represented (segments): Hebrew language vs. Russian language users. Retrieved on 10/5/18.

Figure 30 shows a geographic location report of GA for the use of the Kolzchut website during the 15/10/17-14/11/17 time period. As the geographical location is set by the users' IP addresses, this report is not very useful in Israel.

Figure 30: Google Analytics geographic location report of the Kolzchut website[27]



Challenges in the Visualization of Trace data in the Digital Divide Context

Visualizing trace data is complicated by its nature. Items might present various types of entities (e.g. users, sessions, hits, conversations) that can be grouped in different ways (e.g. subsets of unique users by specific gender, age group, language and combinations of these groups). Attributes of the entities might be numeric or textual. Data is multidimensional, as items might be characterized by several attributes (e.g. a conversation in the internet could be characterized by the media channel used, the specific website in use or by the frequency of specific keywords). The internet content usage categories contribute to its dimensionality as well. Trace data, as being time-oriented data, could to be explored in various time granularities (e.g. days, months or years) and different time arrangement (linear or cyclic). Graphically displaying data by commonly used instruments such as line and bar charts might have an advantage due to their familiarity on the one hand, but on the other hand they might be too "heavy" or "crowded" due to the scope of the data. As the main goal of information visualization is to simplify the information for the user (Munzner, 2014), it seems that innovative visualization methods could be helpful, despite their unfamiliarity.

Careful attention to data abstraction and especially to normalization issues must be taken. For example, some tools provide metric reports segmented by demographic attributes such as gender and age (e.g. GA and SimilarWeb tools). However, such comparison will

[27] GA users' maps. Metric: unique users. Population represented: Hebrew vs. Arabic vs. Russian language users. Retrieved on 10/5/18.

not make much sense unless the data is normalized to reflect the general population distribution (excluding the youngest age cohort). For example, Figure 24 presents interactive chart from a SimilarWeb report, showing the gender and age distribution of unique users in a website designed for senior citizens (motke.co.il). While examining the difference between male and female distribution is possible, making inferences about the difference in the distribution of the two age groups (e.g. low use among age group 65+) is problematic since the general age distribution of the population is not reflected in the chart.

As trace data include textual items (e.g. conversations, search keywords and landing pages titles), providing texts visualizations would be useful.

There are many directions for future work with regards to the visualization of trace data in general, and more specifically with respect to the visualization of the digital divide. For example, there are open research questions regarding faceted analytical displays. How many displays can be combined on the same screen? How many different types of charts can be efficiently shown and understood? Do the answers for these questions depend on the type of metrics presented? Another direction would be to develop and evaluate innovative visualization methods that will represent trace data for the purpose of digital divide evaluation.

Chapter 6: Summary and Conclusions

In this project, an innovative approach for mapping and analyzing the digital divide in Israel was applied using unobtrusive, multi-source digital-trace data. The study employed a wide range of descriptive and quantitative statistical methods (e.g. graphical display of digital gaps, application of various statistical tests and models, formulation of a specially tailored normalized index for the evaluation of gaps in internet content usage) as well as qualitative tools, involving textual analysis of on-line discussions, reflecting on digital gaps.

A triangulation-based approach was used to evaluate and analyze differences in online user-behavior in order to deepen the understanding of the digital divide phenomenon and to construct more robust measurements, allowing evidence-based evaluation of actions. The triangulation approach involved the combination and application of several methods and tools with the specific aim of facilitating the understanding of the digital divide phenomenon. This methodology was demonstrated by a case-study that investigated and analyzed digital gaps in the rights realization domain and involved the use of data stories that supplied systematic guidance for researching and understanding these divides. The data-driven stories were subsequently portrayed by data visualization.

The findings of the research pointed out to the existence of digital gaps, as reflected by ***usage volume*** (number of visits/distribution of visits), ***variety*** (the number of different website categories visited by the user) and ***content usage*** (type of on-line activities or content consumed), with the latter category being the most significant in terms of gaps out of the three.

In terms of usage volume, male users were found to exhibit higher usage volume than female users, having on average 33% more visits than female users. Website visits were found to decrease with age, with the 25-35 age group being the most active in terms of internet use intensity. Significant differences in usage volume were also observed between Hebrew, Arabic and Russian speakers. The usage volume among Hebrew speakers was two times larger than Arabic speakers and 2.4 times larger than native Russian speakers. Stark spatial differences in usage volume were found between users from the core region (Tel Aviv District) and the country's periphery (North and South Districts) , with the usage volume characterizing Core residents being five times higher than the one characterizing users from the Periphery.

In terms of Internet content diversity, male users were found to be more diverse than female users with respect to internet content consumption. The diversity level of the 55+ age group was found to be significantly higher than those of all other age groups with the exception of the 45-54 age group. Spatial differences with respect to the diversity level were also found to be significant, with users from the Tel Aviv metropolitan region and the Jerusalem metropolitan region exhibiting the highest diversity levels and statistically differ from users from other regions. The level of diversity was found to rise with education level, where individuals with post-secondary education or higher had a substantially higher diversity level than individuals holding secondary education or lower. Surprisingly, ultra-orthodox users exhibited the highest level of diversity and statistically differ from all other groups. The data also revealed rather small differences in diversity with respect to income levels.

In terms of content usage, the findings of the research reveal a high degree of compatibility between the results of this digital trace exercise and various findings reported in the literature:

- Digital gaps in online behavior between female and male users was found to be substantial in the following content usage categories: Information and Search; Entertainment; Finance; Dating (dominated by males) and Health (dominated by female).

- Substantial age-based differences in online behavior were identified in the following content usage categories: E-mail; Health; On-line shopping (dominated by older age cohorts) and Entertainment (dominated by younger age cohorts).

- Significant education-based differences in online behavior were identified in the following content usage categories: Government and rights realization; News; Work, career, research and education; Finance (dominated by users with higher levels of education) and Entertainment (music, video and gaming etc.); Communication tools, Instant messaging, chat and social networks and Gambling (dominated by users with lower levels of education).

- Significant differences in online behavior between high-income users and low-income users were identified in the following content usage categories: Entertainment (music, video and gaming etc.); Communication tools, Instant messaging, chat and social networks (dominated by users with lower levels of income) and Travel and tourism (dominated by users with higher levels of income).

The findings of the rights realization case study demonstrated the feasibility of digital divide evaluation by the means of trace data triangulation. Overall, five different digital trace data sources were used in the triangulation process: Buzzilla, Ifat dataset, SimilarWeb, Google Analytics and Google Trends. With respect to rights realization divides, the triangulation methodology has revealed the following insights:

- Females tend to be slightly less active than males with respect to the realization of rights. Rights realization decreases with age, whereas young users are the most active in the realization of rights and the activity of older users in this respect is significantly lower.
- The share of users requiring mediation for the realization of their rights (e.g. people who are assisted by others for the search of information) is significantly higher among older populations.
- A discourse concerning minimum wage right realization dominates the employee rights realization domain.
- Temporal similarities in the distribution of rights realization discussion volumes can be observed both over a one-month and a one-year periods.
- Examining digital trace data concerning entitlements facilitates an understanding of the naming procedure (i.e. the ability of a user to provide an accurate name for a specific right, which is required for its realization).
- Facebook constitute the most popular social media channel for rights realization, while Twitter and Blogs are the least popular.
- News media coverage prompt public interest, active involvement and contribute to public discourse in the rights realization domain.

Chapter 7: Limitations, Contributions and Policy Implications

The Limitations of the Research

This research has several limitations. The most notable constraint of the research is that these data track the activity and behavior of internet users and exclude the activity of non-user populations. Thus, any digital gaps that may be identified will only reflect the divides that exist among on-line users. In this respect, it may very well be that this bias is not very significant as recent studies show that the concept of the digital divide has shifted from denoting gaps in access to information technologies to highlighting differences in the on-line content consumed by users.

A second limitation relates to the temporal dimension of the research. This concerns two datasets that are specifically important to the study of the socio-demographic dimension of the digital gaps - the Ifat Panel dataset and the SimilarWeb Learning Set. Due to the fact that these two datasets were provided to SNI either on a gratis basis or for a relatively small, symbolic fee, they only include short-term data (one month in the case of Ifat Panel dataset and one year in the case of Similar Web Learning Set). Due to this limitation, these two types of digital trace artefacts provided a rather static picture of the digital divide.

A third constraint relates to methodological aspects of data selection and data design. SNI gained access to "off-the-shelf" datasets and digital platforms, of which some were rather raw in their nature (e.g. Ifat Panel dataset), some took a more structured form (e.g. SimilarWeb Learning Set) while others were closed and predefined (e.g. Buzzilla, Google Analytics, Google Trends, SimilarWeb online). As a result, the data collection process (e.g. selection of variables and surveyed websites) and the sampling procedure were solely determined by the entities who formulated the datasets and platforms, giving SNI researchers little flexibility and ability to interfere and re-structure the data.

A fourth limitation concerns the ability to make accurate comparisons between the various trace data sources due to the challenge of multiple sources integration (e.g. inconsistency in the representation of all device types - smartphones, tablet and desktop). While some sources reflect desktop use only (e.g. Ifat panel data), other sources reflect all device use (e.g. Buzzilla). Different time granularities and time-range extractions are additional examples for multiple-source integration limitation (e.g. the time-range criterion in GA is defined by specific dates whereas in SimilarWeb it is defined by a full month period).

Finally, a large dose of modesty should be used in making inferences and drawing conclusions about the digital divide based on a relatively limited collage of digital trace data, as this type of data is virtually infinite, dynamic and constantly evolving.

The Contributions of the Research

This project has provided a proof of concept and important insights regarding the use of digital trace data in the study of on-line user behavior in general and the characterization of the digital divide in particular and as to the ability of unobtrusive tools to replace self-report methods in this task.

The project offers several novel methodological, theoretical and practical contributions to the study of digital divide. To the best of our knowledge, no research to this date has offered a comprehensive methodological framework for measuring and analyzing the socio-demographic aspects of the digital divide within a country using multi-source digital trace data. In this respect, Segev and Ahituv (2010) have come the closest in their seminal work on the development of a new methodology and metrics to examine and assess the digital divide in information searches conducted on the Google and Yahoo platforms. Their study however, focused on divides between countries, did not take into account the socio-economic characteristics of on-line users and did not use multi-source digital trace data. Thus, addressing these aspects in the framework of this research constitutes a clear and significant methodological contribution to the body of knowledge.

The project also offers several practical novelties, crucial for the work of policy and decision makers. The ability of digital trace techniques to capture data “on demand”, to facilitate benchmarking and to present relevant indices and measurements at a detailed temporal, spatial and content usage levels provides decision makers and stakeholders the ability to receive high resolution data at sectoral levels. This type of high-end resolution is missing for example in the National ICT index, despite being an extremely groundbreaking enterprise due to its use of advanced CI techniques. The use of visualizations in the framework of the research and the use of data-driven stories could contribute in raising stakeholder’s interest, promoting transparency, understanding and trust in the data, and is of high relevance to the media and the public at large.

Policy Implications

The findings and outputs of this project offer numerous practical contributions and insights, which may benefit policy and decision makers, as well as the research community at large.

The research has clearly demonstrated that the problem of data anonymity and data confidentiality could be responsibly and safely addressed in the framework of big data and digital trace research, without conceding any harmful information that may have an impact on the privacy of on-line users. In this regard, an enormous gap exists between the above insights regarding the safeguarding of data anonymity on the one hand, and the willingness to grant access to this type of data on behalf of commercial (e.g. private companies) and especially government actors, on the other hand. This gap is expected to grow even wider with the implementation of the General Data Protection Regulation (GDPR) which came into force on May 25th, 2018. The GDPR is concerned with data protection and privacy for all individuals within the European Union (EU) and the European Economic Area (EEA) and addresses the export of personal data outside the EU and EEA areas. Under the GDPR, the data controller must implement measures which meet the principles of data protection by design. A failure to do so might lead to severe sanctions and heavy fines.

The findings and outputs of this digital trace data research supply government actors in Israel valuable insights for the formulation of public policy. Yet, as experienced in this project, there is an increasing difficulty in obtaining access to digital trace data from government actors (e.g. data reflecting online behavior of users in gov.il website and other government websites). This reluctance on behalf of private and government actors to share or make available digital data (either free of charge or for a fee) is largely due to the existence of legal ambiguities and unwillingness on behalf the various data owners to deal with potential problems that might arise from granting this use. These trends cast a dark shadow on the future of academic research concerning digital traces and thus should be of a keen interest to the regulator and to policy makers.

In this regard, we propose the Ministry of Science and Technology the following:

To be active in formulating a protocol that will define and regulate the release and use of digital trace data for research purposes*. *This protocol should set clear guidelines for:

- Data collection and data mining from on-line sources.

- The anonymization (or aggregation) of personal information on behalf of the data owner.
- Accepted practice and procedures for data processing, cross referencing and consolidation of digital trace data and survey data from multiple sources.
- Guidance regarding the presentation of the data (on behalf of the researcher).
- Transparency and third-party use.
- The construction and maintenance of digital trace repositories (with or through entities such as the National Library or the Israel State Archives-ISA) that will ensure that anonymized longitudinal digital trace data will be retained and available for the use of the research community.
- The penalties that might be imposed on the researcher in case of breaching the contract terms etc.

The findings of this exercise make it possible to infer policy conclusions with regards to the mitigation of digital gaps in several content domains. These mostly involve on-line activities pertaining to every-day life such as health, finance and real-estate, rights realization, E-government and gambling. In this respect, ***we recommend the relevant government offices, service and data providers of on-line platforms (e.g. banks, e-health and municipal service providers, universities, etc.) and social society actors (e.g. NGOs involved in making online information accessible to the public) to***:

- Raise awareness and enhance education, especially among women, young adults, lower income and lower education populations as to the ***importance of on-line financial education and knowledge about the housing market*** (e.g. searching and accruing information on pension and provident funds, conducting on-line bank transactions and investments, acquiring mortgages, etc.).

- Raise awareness and enhance education, especially among men, young adults, lower income and ultra-orthodox populations as to the ***benefits of using and conducting e-health activities*** (e.g. conducting e-health activities such as making doctor appointments, requesting on-line prescriptions, searching for information on health-related issues).

- Raise awareness and enhance education, especially among older adults, lower income and lower education populations as to the benefits of using ***e-gov and on-line municipal services*** (e.g. use of various e-gov services such as passport renewal, information on income taxes, paying local taxes such as water and property tax/arnona, land and property registration etc.).

- To track user behavior in various rights realization and e-government websites in order to identify ***popular and frequently searched rights*** that interest the public on the one hand, and on the other hand, to identify the ***rights which are seldomly searched and only partially realized by their eligible and potential beneficiaries***. The tracking of these websites should be aimed at identifying temporal trends in user behavior pertaining to these two types of rights.

- Raise awareness and enhance education, especially among women, young adults, ultra-orthodox and lower education populations as to the importance of searching information and conducting on-line transactions with regards to ***entitlement benefits and rights realization*** (e.g. accessing information on social security benefits, maternity rights and payments, minimum wage, disabilities etc.).

- The research has exposed the importance of defining and using accurate search terms in retrieving information (see the naming story). Google Trends was found to be a useful tool for this purpose. This ability could be of especially high value with regards to the realization of rights, as using accurate search terms and better queries may lead to higher access to information about these rights, thus leading to their realization. We recommend the relevant government actor (National Insurance Institute) to learn about the variant use of each right by its users (understanding user behavior in applying different search terms for the same right), with the specific aim of better customizing relevant websites (e.g. www.btl.gov.il).

- Encourage the use of social networks and blogs (especially among ultra-orthodox population) among government offices in disseminating knowledge and raising public awareness in the domain of ***rights realization***, with the specific aim of targeting deprived populations.

- To conduct continuous monitoring and analysis of the discourse in the rights realization domain, taking place in official websites (Facebook, forums, blogs etc.) of government ministries and agencies, with the specific aim of deepening our understanding as to the needs of the active users, as well as identifying additional audiences who do not participate in this discourse.

- Raise awareness and enhance education, especially among older adults and lower income population as to the importance and benefits ***of e-education and e-learning activities*** (e.g. on-line courses, use and access to on-line infrastructure in universities).
- Raise awareness and address the problem of increased on-line ***gambling activity*** especially among men, young adults, lower income and lower education populations.

In addition, important ***methodological and procedural lessons*** could be learnt from this research which could be of value to the ***research community***:

- A key contribution of this research was the demonstration of a triangulation methodology that provided a multi-faceted view of the digital divide (with zoom-in on gaps in rights realization), integrating numerous digital trace data sources and methods for the purpose of enhancing the understanding and reliability of the data and the investigated phenomena. This initial work has stressed the need for further developing empirical tools for consolidating insights derived from quantitative data (e.g. web visits) and perceptions and sentiment derived from the analysis of textual and lexical data (e.g. analysis of web discussion in social networks and blogs). We recommend the research community to take a step at this direction as it may significantly contribute to the understanding of investigated phenomena (e.g. why do digital divides exist).
- As digital divides are shifting in practice from gaps in access to information technologies to gaps in the range and distribution of on-line content consumed by users, the formulation of a taxonomy for classifying and organizing internet content is becoming ever more imperative. This research has introduced a manual procedure for the categorization of content usage due to the relatively modest scope of our trace data. However, such procedure is not applicable in the case of truly large corpus data. Thus, machine learning and AI techniques should be developed and applied in this domain to facilitate automatic content classification of websites. This task is related to Natural Language Processing (NLP) which is especially challenging with regards to the Hebrew language. A research effort in this direction should be mobilized by the research community.

- The research has demonstrated the contribution of consolidating internet panels with digital traces in deepening our understanding of online user behavior and the digital divide. However, it is important to remember that although overt behavior was tracked in the framework of this research, it was based on a small subset of the user population, thus making it prone to sampling bias. This selection bias may reflect unrepresentative selection of the population (e.g. its socio-demographic composition) or the content consumed by the users (e.g. website selection). More specific to the digital divide, the panel method may not provide adequate coverage of less frequently visited websites which may be important to the understanding of specific aspects of the phenomena. Working on digital traces based on internet panels also requires placing much attention on data cleansing (e.g. deleting the extensive reference to panel websites themselves) and outlier observations. More effort on behalf the research community should be directed in tackling these problems and questions.

- As visualizing the digital divide is highly important, developing design guidelines for digital divide visualization based on trace data is essential. Trace data in the digital divide context is characterized by high volume of multi-dimensional, time-oriented data. Another aspect to consider is the challenge of organization and integration of multiple sources of data. A research effort in this direction should be mobilized by the research community.

- To encourage the cooperation of the academia with commercial companies engaged in the monitoring and analysis of digital trace data, as well as with NGOs focusing on accessing on-line information and knowledge to the public (e.g. rights realization) for the purpose of conducting joint research projects.

The last five recommendations also constitute our reflections for further research in the evaluation of digital trace data in general and the study of digital divide in particular.

List of References

Abbey, R., & Hyde, S. (2009). No country for older people? Age and the digital divide. Journal of information, Communication and Ethics in Society,7(4), 225-242.

Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). *Visualization of time-oriented data.* Springer Science & Business Media.

Akca, H., Sayili, M., & Esengun, K. (2007). Challenge of rural people to reduce digital divide in the globalized world: Theory and practice. Government Information Quarterly, 24(2), 404-413.

Alam, S. S., Abdullah, Z., & Ahsan, N. (2009). Cybercafé usage in Malaysia: An exploratory study. Journal of Internet Banking and Commerce,14(1), 1.

Albo, Y., Lanir, J., & Rafaeli, S. (2017). A Conceptual Framework for Visualizing Composite Indicators. *Social Indicators Research*, 1-30.

Albo, Y., Lanir, J., Bak, P., & Rafaeli, S. (2016). Off the radar: Comparative evaluation of radial visualization solutions for composite indicators. *IEEE transactions on visualization and computer graphics*, *22*(1), 569-578.

Albo, Y., Lanir, J., Bak, P., & Rafaeli, S. (2016, June). Static vs. Dynamic Time Mapping in Radial Composite Indicator Visualization. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (pp. 264-271). ACM.

Avidar, M. 2009. Digital and national divides in Israel. MA thesis. University of Haifa: Department of sociology and anthropology.

Barak-Erez, D., & Gross, A. (Eds.). (2007). *Exploring social rights: Between theory and practice*. Bloomsbury Publishing.

Barrantes, R., & Galperin, H. (2008). Can the poor afford mobile telephony? Evidence from Latin America. Telecommunications Policy, 32(8), 521-530.

Barzilai-Nahon, K. (2006). Gaps and bits: Conceptualizing measurements for digital divide/s. The information society, 22(5), 269-278.

Barzilai-Nahon, K., Rafaeli, S., & Ahituv, N. (2004). Measuring Gaps in Cyberspace: Constructing a comprehensive digital divide index. In Workshop on Measuring the Information Society, the conference of Internet Research (Vol. 5).

Bell, P., Reddy, P., & Rainie, L. (2004). Rural areas and the Internet. Washington, DC: Pew Internet & American Life Project, 7-37.

Benish, A., David, L. (2017) על (אי-)מיצוי זכויות חברתיות זכות הגישה למנהל במדינת הרווחה: וחובת ההנגשה של החקיקה החברתית. Available at http://weblaw.haifa.ac.il/he/Journals/lawGov/Volume19/Avishai%20Benish_Liron%20David.pdf

Beyer, M. A., & Laney, D. (2012). The importance of ‘big data’: a definition. *Stamford, CT: Gartner*, 2014-2018.

Bezeq, (2017). Bezeq Internet Report: 2017 החיים בעולם הדיגיטלי: דו"ח האינטרנט של בזק. Available at https://www.bezeq.co.il/media/PDF/internetreport_2017.pdf

Blank, G., & Groselj, D. (2014). Dimensions of Internet use: amount, variety, and types. *Information, Communication & Society*, *17*(4), 417-435.

Bonfadelli, H. (2002). The Internet and knowledge gaps: A theoretical and empirical investigation. *European Journal of communication*, *17*(1), 65-84.

Brabazon, T. (2010). Unobtrusive Research Methods. From Practising Media Research, Week 2 [Podcast].

Bruno, G., Esposito, E., Genovese, A., & Gwebu, K. L. (2010). A critical analysis of current indexes for digital divide measurement. The Information Society, 27(1), 16-28.

Callegaro, M., & Yang, Y. (2018). The Role of Surveys in the Era of “Big Data”. In *The Palgrave Handbook of Survey Research*(pp. 175-192). Palgrave Macmillan, Cham.

CBS, (2016). Central Bureau of Statistics Israel annual report: אוכלוסיה לפי גיל. Available at http://www.cbs.gov.il/reader/cw_usr_view_SHTML?ID=803

Chakraborty, J., & Bosman, M. M. (2002). Race, income, and home PC ownership: A regional analysis of the digital divide. Race and Society, 5(2), 163-177.

Charleer, S., Klerkx, J., Duval, E., De Laet, T., & Verbert, K. (2016, September). Creating effective learning analytics dashboards: Lessons learnt. In European Conference on Technology Enhanced Learning (pp. 42-56). Springer, Cham.

Chen, W., & Wellman, B. (2003). Charting and bridging digital divides: comparing socio-economic, gender, life stage, and rural-urban Internet access and use in eight countries. In W. Dutton, B. Kahin, R. O’Callaghan and A. Wyckoff (Eds.), Transforming Enterprise (pp. 467-97). Cambridge, MA: MIT Press.

Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007). An introduction to ‘benefit of the doubt’ composite indicators. Social Indicators Research, 82(1), 111–145.

Cornfield, M., & Rainie, L. (2003). Untuned keyboards: Online campaigners, citizens, and portals in the 2002 elections. PEW Internet and American Life.

Cruz-Jesus, F., Oliveira, T., & Bacao, F. (2012). Digital divide across the European Union. Information & Management, 49(6), 278-291.

Cukier, J. 2011. Can Data Visualization Help Build Democracy? *XRDS: Crossroads, ACM Magazine for Students*. 18, 2: 26-30.

Cullen, R. (2001). Addressing the digital divide. Online information review,25(5), 311-320.
Denzin, N. K. (1973). *The research act: A theoretical introduction to sociological methods*. Transaction publishers.

Dickson, G. W., Senn, J. A., & Chervany, N. L. (1977). Research in Management Information Systems: The Minnesota Experiments. *Management Science*, 23(9), 913-923.

DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). From unequal access to differentiated use: A literature review and agenda for research on digital inequality. Social inequality, 355-400.

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, *17*(2), 245-260.

Dutton, W., Helsper, E., & Gerber, M. (2011). The internet in Britain in 2011.

Ebo, B. L. (1998). Cyberghetto or cybertopia?: race, class, and gender on the Internet. Santa Barbara: Praeger.

Elting, L. S., Martin, C. G., Cantor, S. B., & Rubenstein, E. B. (1999). Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *Bmj*, *318*(7197), 1527-1531.

Enoch, Y., & Soker, Z. (2006). Age, gender, ethnicity and the digital divide: university students' use of web-based instruction. Open Learning, 21(2), 99-110.

Felstiner, W. L., Abel, R. L., & Sarat, A. (1980). The Emergence and Transformation of Disputes: Naming, Blaming, Claiming... *Law and society review*, 631-654.

Few, S., & Edge, P. (2007). Dashboard confusion revisited. *Perceptual Edge*, 1-6.

Foulger D. 2001. Seven Bridges Over the Global Digital Divide, IAMCR & ICA Symposium on Digital Divide.

Fox, S., & Madden, M. (2006). Generations online (demographic report). Pew Internet\& American Life Project.

Fuchs, C. (2009). The role of income inequality in a multivariate cross-national analysis of the digital divide. Social Science Computer Review, 27(1), 41–58

Ganayem, A., S. Rafaeli, & F. Azaiza. 2009. Digital divide: Internet usage in the Arab sector in Israel. Megamot, 36: 164-196. (Hebrew).

Glass, V., & Stefanova, S. K. (2010). An empirical study of broadband diffusion in rural America. Journal of Regulatory Economics, 38(1), 70-85.

Graham, J. W., Collins, N. L., Donaldson, S. I., & Hansen, W. B. (1993). Understanding and controlling for response bias: Confirmatory factor analysis of multitrait-multimethod data. *Psychometric methodology*, 585-590.

Halpin, H., Robu, V., & Shepherd, H. (2007, May). The complex dynamics of collaborative tagging. In Proceedings of the 16th international conference on World Wide Web (pp. 211-220). ACM.

Hampton, K. N. (2017). Studying the Digital: Directions and Challenges for Digital Methods. *Annual Review of Sociology*, *43*, 167-188.

Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults' use of the Internet. *Communication research*, *35*(5), 602-621.

Helsper, E. J., & Galácz, A. (2009). Understanding the links between social and digital exclusion in Europe. *World Wide Internet: Changing Societies, Economies & Cultures, University of Macau, Taipa, China*, 146-178.

Hitt, L., & Tambe, P. (2007). Broadband adoption and content consumption. Information Economics and Policy, 19(3), 362-378.

Hoffman, D. L., Novak, T. P., & Schlosser, A. (2000). The evolution of the digital divide: How gaps in Internet access may impact electronic commerce. Journal of Computer-Mediated Communication, 5(3).

Hoskin, R. (2012). The dangers of self-report. *London: British Science Association*.

Howard, P. E., Raine, L., & Jones, S. (2001). Access, Civic Involvement, and Social Interaction. *American Behavioral Scientist*, *45*(3), 382-404.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. Journal of the Association for Information Systems, 12(12), 767.

Jackson, L. A., Ervin, K. S., Gardner, P. D., & Schmitt, N. (2001). Gender and the Internet: Women communicating and men searching. Sex roles, 44(5-6), 363-379.

James, J., & Versteeg, M. (2007). Mobile phones in Africa: how much do we really know?. Social indicators research, 84(1), 117-126.

Jones, S., & Fox, S. (2009). Pew internet project data memo. *Pew internet & American life project*.

Jones, Q., & Rafaeli, S. (2000, January). What do virtual" Tells" tell? Placing cybersociety research into a hierarchy of social explanation. In *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on* (pp. 10-pp). IEEE.

Katz, J., & Aspden, P. (1997). Motivations for and barriers to Internet usage: Results of a national public opinion survey. Internet Research, 7(3), 170-188.

Kellehear, A. (1993). The Unobtrusive Researcher. Sydney, Australia: Allen and Unwin.

Kennedy, T., Wellman, B., & Klement, K. (2003). Gendering the digital divide. It & Society, 1(5), 72-96.

Kim, Y. C., Jung, J. Y., & Ball-Rokeach, S. J. (2007). Ethnicity, place, and communication technology: Effects of ethnicity on multi-dimensional internet connectedness. Information Technology & People, 20(3), 282-303.

King, G., Schneer, B., & White, A. (2017). How the news media activate public expression and influence national agendas. *Science*, *358*(6364), 776-780.

Kourtit, K., & Nijkamp, P. (2018). Big data dashboards as smart decision support tools for i-cities–An experiment on stockholm. *Land Use Policy*, *71*, 24-35.

Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. Journal of marketing research, 39(1), 61-72.

Laflaquiere, J. (2009). Conception de système à base de traces numériques dans les environnements informatiques documentaires (Doctoral dissertation, Troyes).

Lim, J. (2002). East Asia in the Information Economy: Opportunities and challenges. Info, 4(5). 56 – 63.

Losh, S. C. (2004). Gender, educational, and occupational digital gaps 1983-2002. Social Science Computer Review, 22(2), 152-166.

Madden, M. (2003). America’s online pursuits. Washington, DC: PEW Internet and American Life Project.

McLaren, J., & Zappala, G. (2002). The 'digital divide' among financially disadvantaged families in Australia. First Monday, 7(11).

Menou, M. J., & Taylor, R. D. (2006). A “grand challenge”: Measuring information societies. The Information Society, 22(5), 261-267.

Middleton, K. L., & Chambers, V. (2010). Approaching digital equity: is wifi the new leveler?. *Information Technology & People*, *23*(1), 4-22.

Moon, J., Park, J., Jung, G. H., & Choe, Y. C. (2010). The impact of IT use on migration intentions in rural communities. Technological Forecasting and Social Change, 77(8), 1401-1411.

Munzner, T. (2014). *Visualization analysis and design*. CRC press.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). Handbook on constructing composite indicators.

Noce, A. A., & McKeown, L. (2008). A new benchmark for Internet use: A logistic modeling of factors influencing Internet use in Canada, 2005. Government Information Quarterly, 25, 462–476

O'Brien, M. (2010). Unobtrusive research methods-An interpretative essay.

OECD. Publishing, & Organisation for Economic Co-operation and Development. (2011). *OECD Guide to Measuring the Information Society 2011*. Organisation for Economic Co-operation and Development.

OECD/DSTI. 2001. Understanding the digital divide. OECD papers.

Orviska, M., & Hudson, J. (2009). Dividing or uniting Europe? Internet usage in the EU. Information Economics and Policy, 21(4), 279-290.

Park, S. (2009, March). Concentration of internet usage and its relation to exposure to negative content: Does the gender gap differ among adults and adolescents?. In *Women's Studies International Forum* (Vol. 32, No. 2, pp. 98-107). Pergamon.

Park, E. A., & Jayakar, K. (2010). Patterns of E-rate funding to school districts: An eight state comparison. info, 12(3), 46-58.

Peter, J., & Valkenburg, P. M. (2006). Adolescents’ internet use: Testing the “disappearing digital divide” versus the “emerging digital differentiation” approach. Poetics, 34(4), 293-305.

Petrović, M., Bojković, N., Anić, I., & Petrović, D. (2012). Benchmarking the digital divide using a multi-level outranking framework: Evidence from EBRD countries of operation. Government Information Quarterly, 29(4), 597-607.

Pook, L. A., & Pence, N. E. (2004). Evaluation of information infrastructures and social development among the Visegrad-Four countries of Central Europe. Journal of Global Information Management (JGIM), 12(2), 63-83.

Prieger, J. E., & Hu, W. M. (2008). The broadband digital divide and the nexus of race, competition, and quality. Information economics and Policy,20(2), 150-167.

Pulse, U. G. (2016). Integrating Big Data into the Monitoring and Evaluation of Development Programmes.

Rafaeli, S., Albo, Y. and Shiti, I. (2013) Israel National ICT Index - Research progress report on the creation of ICT Index to promote international technology and Internet use in Israel. Haifa: The Center for Internet Research. (in Hebrew).

Rice, R. E., & Katz, J. E. (2003). Comparing internet and mobile phone usage: digital divides of usage, adoption, and dropouts. Telecommunications Policy, 27(8), 597-623.

Riche, N. H., Hurter, C., Diakopoulos, N., & Carpendale, S. (Eds.). (2018). *Data-Driven Storytelling*. CRC Press.

Robinson, J. P., DiMaggio, P., & Hargittai, E. (2003). New social survey perspectives on the digital divide. It & Society, 1(5), 1-22.

Rudder, C. (2014). *Dataclysm: Who we are (when we think no one's looking)*. Random House Canada.

Savage, S. J., & Waldman, D. M. (2009). Ability, location and household demand for Internet bandwidth. International Journal of Industrial Organization, 27(2), 166-174.

Schleife, K. (2010). What really matters: Regional versus individual determinants of the digital divide in Germany. Research Policy, 39(1), 173-185.

Schober, M. F., Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F. G. (2016). Social media analyses for social measurement. *Public opinion quarterly*, *80*(1), 180-211.

Sciadas, G. (2004, September). International benchmarking for the information society. In *ITU-KADO digital bridges symposium*.

Segev, E., & Ahituv, N. (2010). Popular searches in Google and Yahoo!: A “Digital Divide” in information uses?. *The Information Society*, *26*(1), 17-37.

Shirazi, F., Ngwenyama, O., & Morawczynski, O. (2010). ICT expansion and the digital divide in democratic freedoms: An analysis of the impact of ICT expansion, education and ICT filtering on democracy. Telematics and Informatics, 27(1), 21-31.

Schumacher, P., & Morahan-Martin, J. (2001). Gender, Internet and computer attitudes and experiences. *Computers in human behavior*, *17*(1), 95-110.

Siliverstovs, B., & Wochner, D. S. (2018). Google Trends and reality: Do the proportions match?. *Journal of Economic Behavior & Organization*, *145*(C), 1-23.

Stocking, G. and Matsa, K. E. (2017). Using Google Trends data for research? Here are 6 questions to ask. Pew Research Center.

Subrahmanyam, K., Kraut, R., Greenfield, P., & Gross, E. (2001). New forms of electronic media. Handbook of children and the media, 73-99.

Tien, F. F., & Fu, T. T. (2008). The correlates of the digital divide and their impact on college student learning. Computers & Education, 50(1), 421-436.

Trauth, E. M. (2002). Odd girl out: an individual differences perspective on women in the IT profession. Information Technology & People, 15(2), 98-118.

UN Global Pulse (2015), Mining Citizen Feedback Data for Enhanced Local Government Decision-Making. Global Pulse Project Series, (16).

Van Deursen, A. J., & Van Dijk, J. A. (2014). The digital divide shifts to differences in usage. *New media & society*, *16*(3), 507-526.

Van Dijk, J. A. (2006). Digital divide research, achievements and shortcomings. Poetics, 34(4), 221-235.

Vehovar, V., Sicherl, P., Hüsing, T., & Dolnicar, V. (2006). Methodological challenges of digital divide measurements. The information society, 22(5), 279-290.

Viégas, F. B., Boyd, D., Nguyen, D. H., Potter, J., & Donath, J. (2004, January). Digital artifacts for remembering and storytelling: Posthistory and social network fragments. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on* (pp. 10-pp). IEEE.

Webb, Eugene J. 2000. Unobtrusive Measures. Rev. ed, Sage classics series. Thousand Oaks, Calif.: Sage Publications.

Weiss-Gal, I., & Gal, J. (2009). Realizing rights in social work. *Social Service Review*, *83*(2), 267-291.

Wilbon, A. D. (2003). Shrinking the digital divide: the moderating role of technology environments. Technology in Society, 25(1), 83-97.

Xiong, R., & Donath, J. (1999, November). PeopleGarden: creating data portraits for users. In *Proceedings of the 12th annual ACM symposium on User interface software and technology* (pp. 37-44). ACM.

Zilka, G. (2012). Reducing the digital gap among underserved populations in Israel. Ma'of u-Ma'aseh Teaching and Learning in the Internet Era 14. Achva Academic College. 101-138. (Hebrew)

Zillien, N., & Hargittai, E. (2009). Digital distinction: Status-specific types of internet usage. *Social Science Quarterly*, *90*(2), 274-291.

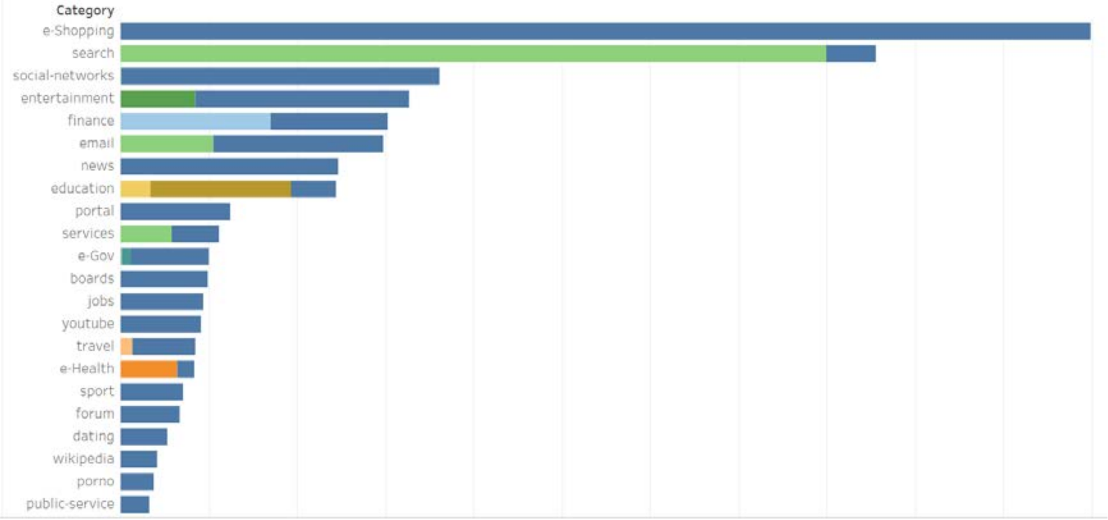
Digital Israel (2017). The National Digital Plan of the Government of Israel. המיזם הלאומי ישראל דיגיטלית, התכונית הדיגיטלית הלאומית של ממשלת ישראל. Available at: https://www.gov.il/BlobFolder/news/digital_israel_strategy/he/Digital_Israel_0.pdf

State Comptroller of Israel Annual Report (2015). Social right non realization. דו"ח שנתי 65ג, אי-מיצוי של זכויות חברתיות. Available at: http://www.mevaker.gov.il/he/Reports/Report_290/7cf0fcb2-918b-417a-94f2-75032d49d01c/65C-101-ver-3.docx

Annex 1: Ifat categorization methodology

- First, I describe steps on detail. A short paragraph can be found below.
- We produced a “Clean URL” file in which records referring to same sites were merged. Number of entries to each website was documented (naming “Frequency” field). As a result, the 2,000,000 original records were reduced to 41,518 records representing all panel’s activity.
- Records were sorted by number of website’s instances in the “Full URL” file (i.e. by “Frequency” field).
- We coded records from top frequent in descending order, to code as much activity as possible.
- Each record was manually coded to a single category. Category type was decided by manually entering each website and choosing the major theme which best characterized it. Two persons executed the coding process separately.
- We constructed the list of the categories “on the fly”, balancing the need to express the diversity of activity vs. the need to reduce categories to minimum. 29 categories were defined.
- We added sub-categories to some of the categories (e.g. “Institutions” in “e-Health” and “education”).
- We used string functions to quickly find all records that met a criterion (e.g. “bank” string for the “finance” category, “doctor” string for e-Health).
- Overall, 90.45% of activity was coded, made by 18.3% of the sites (about 7,600 records were categorized).
- 9.69% of the “Clean URL” records found to be “junk” records, as they presented panel activity (e.g. connecting to various panels’ sites) which is useless for research purposes.

The following figure shows the top 22 categories.



Annex 2: Ifat content usage categories and selected websites belonging to each category

Category	Selected websites belonging to the category (highest frequencies)
Boards	yad2.co.il; agora.co.il; winwin.co.il; homeless.co.il
Communication	hot.net.il; hotmobile.co.il; golantelecom.co.il; 012mobile.co.il; pelephone.co.il; cellcom.co.il; kamaze.co.il; biz.partner.co.il
Dating	okcupid.com; mybf.co.il; dating.atraf.co.il; onlyu.co.il; rusdate.co.il
E-Gov	edu.gov.il; misim.gov.il; ecom.gov.il; gov.il; apot.justice.gov.il; cms.education.gov.il; taxes.gov.il; mitgaisim.idf.il
E-Health	e-services.clalit.org.il; online.maccabi4u.co.il; meuhedet.co.il; clalit.co.il; maccabi4u.co.il; ncbi.nlm.nih.gov; serguide.maccabi4u.co.il; camoni.co.il
E-Shopping	ebay.com; aliexpress.com; he.aliexpress.com; amazon.com; next.co.il; zap.co.il
Education	sheilta.apps.openu.ac.il; manbasnet.education.gov.il; ebag.cet.ac.il; moodle.technion.ac.il; mw5.haifa.ac.il; opal.openu.ac.il; moodle2.cs.huji.ac.il; moodle.sapir.ac.il ;xmail.weizmann.ac.il
Email	mail.google.com; mail.walla.co.il; api-mail.walla.co.il; friends.walla.co.il; baba-mail.co.il; outlook.live.com; mail.partner.net.il; newmail.012.net.il ;012mail.net
Entertainment	mako.co.il; travian.co.il; reshet.tv; sparo.live; netflix.com; kan.org.il; subscenter.info; 10tv.nana10.co.il ;seret.co.il
Finance	login.bankhapoalim.co.il; paypal.com; online.fibi.co.il; talniri.co.il; mizrahi-tefahot.co.il; online.leumi-card.co.il; leumi-card.co.il; services.cal-online.co.il ;
Forum	fxp.co.il; tapuz.co.il; prog.co.il; bhol.co.il; stips.co.il; sat-israel.co.il; bizportal.co.il; stackoverflow.com ;rotter.net
Gambling	pais.co.il; freelotto.com; multicoinfaucet.com; coinbulb.com; pokerland-il.com; lottosheli.co.il; hagralot.info;
Jobs	jobmaster.co.il; drushim.co.il; alljobs.co.il; il.indeed.com; jobnet.co.il; taasuka.gov.il; careers.mobileye.com; ekoclix.com ;clixunion.com
Kids	meirkids.co.il; hop.co.il; kizi.com; yo-yoo.co.il; games.yo-yoo.co.il; crm.meirkids.co.il; g.modelsworld.yo-yoo.co.il; play.mikmak.co.il ;popy.co.il
News	ynet.co.il; rotter.net; globes.co.il; inn.co.il; kikar.co.il; haaretz.co.il; calcalist.co.il; msn.com ;news.walla.co.il
Parcel-Service	17track.net; fcx.co.il; zig-zag.co.il; tracking.i-parcel.com; yaballe.com; webshipping3.dhl.com; fedex.com; dhl.co.il ;ups.com
Adult	Various pornographic websites
Portal	walla.co.il; hidabroot.org; start.co.il; yahoo.com; yehadot.co.il; kafe.co.il
Preservation	mechon-mamre.org; myheritage.co.il; wow.co.il; lupa.co.il; albume.co.il; editor.albume.co.il; zooma.co.il
Public-Service	israelpost.co.il; iec.co.il; accuweather.com; israelweather.co.il; mypost.israelpost.co.il; postil.wizsupport.com; mas23.bezeqint.net; perach.org.il
Real-Estate	madlan.co.il; bvd.co.il; goiconnect.com; cloud.dekel.co.il; remaxfocus.co.il; remax-israel.com; remax.co.il; mls.nadlanone.co.il ;nadlan.what2do.co.il
Rights	lawguide.co.il; btl.gov.il; kolzchut.org.il; ps.btl.gov.il; bankruptcy.what2do.co.il; b2b.btl.gov.il; what2do.co.il; criminal.what2do.co.il ;ovdim.org.il
Search	google.co.il; google.com; search.clearch.org; safesearch.top; d.co.il; search.ask.com; bing.com; google.ru ;scholar.google.co.il
Services	docs.google.com; drive.google.com; calendar.google.com; sites.google.com; login.microsoftonline.com; eu6.salesforce.com; dropbox.com; support.google.com
Social-Networks	facebook.com; web.whatsapp.com; linkedin.com; twitter.com; instagram.com; pinterest.com
Sport	one.co.il; sport5.co.il; sports.walla.co.il; bvl.org.il; winner.co.il; vod.sport5.co.il; football.org.il; live.nivdal.win ;doublepass.sport5.co.il
Translation	translate.google.co.il; morfix.co.il; wooordhunt.ru; translate.google.com; milog.co.il; almaany.com; lyricstranslate.com; translate.google.it ;translate.google.ru
Transportation	bus.gov.il; rail.co.il; egged.co.il; waze.com; bus.co.il; ravkavonline.co.il; dan.co.il
Travel	booking.com; lametayel.co.il; travelist.co.il; elal.com; issta.co.il; tripadvisor.co.il; israir.co.il; wizzair.com ;united.com
Wikipedia	he.wikipedia.org; en.wikipedia.org; ru.wikipedia.org; he.wikisource.org; upload.wikimedia.org; commons.wikimedia.org; ar.wikipedia.org
Youtube	youtube.com

Annex 3: SimilarWeb content usage categories and selected websites belonging to each category

Category	Selected websites belonging to the category (highest frequencies)
Internet and Telecom	kikar.co.il; livejournal.com; dropbox.com; 360.co.il
Arts and Entertainment	9tv.co.il; reshet.tv; mouse.co.il; cinema-city.co.il; yes.co.il; giphy.com; playbuzz.com; zaful.com ;wikia.com
News and Media	walla.co.il; maariv.co.il; themarker.com; tapuz.co.il; israelhayom.co.il; panet.co.il; fxp.co.il; sport5.co.il ;mako.co.il
Shopping	zap.co.il; grouponisrael.co.il; ksp.co.il; super-pharm.co.il; ikea.co.il; zipy.co.il; payngo.co.il; castro.com ;aliexpress.com
Business and Industry	d.co.il; ashdodnet.com
Autos and Vehicles	auto.co.il
Travel	elal.com; booking.com; hotels.com; expedia.com
Health	infomed.co.il; clalit.co.il; iherb.com; doctors.co.il; imaot.co.il
People and Society	hidabroot.org; kipa.co.il; date4dos.co.il
Law and Government	psakdin.co.il; gov.il; justice.gov.il
Finance	mizrahi-tefahot.co.il; investing.com
Career and Education	jobnet.co.il; openu.ac.il
Games	kizi.com
Reference	milog.co.il
Food and Drink	rest.co.il

Annex 4: SimilarWeb content sub-categories and selected websites belonging to each sub-category

Category	Selected websites belonging to the sub-category (highest frequencies)
Online Marketing	360.co.il
TV and Video	9tv.co.il; reshet.tv; yes.co.il; wikia.com
General Merchandise	aliexpress.com; banggood.com
Car Buying	auto.co.il
Business News	themarker.com; bizportal.co.il; calcalist.co.il; globes.co.il
Accommodation and Hotels	booking.com; hotels.com
Magazines and E-Zines	buzzfeed.com
Clothing	castro.com; renuar.co.il
Movies	cinema-city.co.il
Business Services	d.co.il
File Sharing	dropbox.com
Airlines and Airports	elal.com
Tourism	expedia.com
Government	gov.il; justice.gov.il
Coupons	grouponisrael.co.il
Newspapers	maariv.co.il; panet.co.il; haaretz.co.il; nytimes.com; rambler.ru
Religion and Spirituality	hidabroot.org; kipa.co.il
Technology News	hwzone.co.il
Products and Shopping	iherb.com
Investing	investing.com
Jobs and Employment	jobnet.co.il
Online	kizi.com
Consumer Electronics	ksp.co.il
Dictionaries and Encyclopedias	milog.co.il
Banking	mizrahi-tefahot.co.il
Education	openu.ac.il
Law	psakdin.co.il
Restaurants and Delivery	rest.co.il
Sports News	sport5.co.il
Classifieds	zap.co.il

Annex 5: Summary of the rights realization case study stories

Story	Description	Sources used for triangulation	Suggested Implication (for policy makers or *researches)
The gender and age differences story	Indications of gender and age differences on social rights websites use. Lower use among female comparing to male; Lower use among age group 55+.	Gender – SimilarWeb, Ifat Panel, GA; Age – SimilarWeb, GA	Locate specific rights that are salient in non-realization by female or elder people and adopt policies to raise social rights realization among those groups.
The mediators story	Elder people are more often represented by mediators (people who use the internet on behalf of the people who are the subjects of the social right realization) comparing to youngers.	For elderly - Buzzilla, For maternity – Buzzilla, Ifat Panel, GA	Encourage social rights awareness and use by elders or directly relate to their mediators.
The how-much story	Some social rights' volume across sources show slight indication of similarity.	Buzzilla, Ifat Panel, SimilarWeb (Website and Keywords analysis), Google Trends, Google Analytics	*Caution with topics volume comparisons across sources, due to differences in time period covered and device use sampled. Consider comparing topics volume within each source rather than between them. Analyzing proportions of rights' volume across sources might serve as a tool for finetuning the rights' extraction process.
The time range story	One-month rights' volume shows indication of similarity to one-year volume.	Buzzilla, Google Analytics	*As trace data are characterized by high volume, use of a representative time slice for data extraction might be carefully considered.
The naming story	Search keywords analysis might facilitate understanding of the important naming stage (the ability to translate and accurately name a specific benefit).	SimilarWeb, Google Trends	Instill knowledge to the public on the accurate terms of the social rights.
The social media channels story	People discuss social rights on social networks, particularly on Facebook.	Buzzilla, SimilarWeb	Facebook might be useful for raising public awareness to social rights realization.
The "buzz" story	News media activate emergence of public conversations.	Buzzilla (no triangulation)	Articles' publication might be useful for raising public awareness to social rights realization.

Annex 6: “What” aspect - Trace data properties of sources used in this research

Tool / Data source	Right extraction method	Items	Metrics	Device Type	Time period covered	Analyzed websites (access from Israel)
1. Buzzilla	lexical Boolean queries	Conversations’ texts	Number of conversations	Desktop, Mobile, Tablet	15/10/17-14/11/17; 2017	forums, blogs, articles and social networks scanned by Buzzilla
2. Ifat Panel data sub-set	Categories derived by page-title	Landing page titles	Number of visits; Number of visitors	Desktop	15/10/17-14/11/17	Kolzchut & NII
3.1 SimilarWeb-Website Analysis	Categories derived by search terms as sites traffic source	Landing page titles; Search terms referring to analyzed website.	Number of visits; Number of visitors	Desktop only (in demographic reports) or Desktop & Mobile	Oct 17-Nov 17	Kolzchut & NII
3.2 SimilarWeb-Search Keywords Analysis	Categories derived by keyword groups	Search terms	Number of searches		Oct 17-Nov 17	Not relevant
4. Google Trends	Google search terms and combinations of terms.	Search terms	Number of searches	Desktop, Mobile, Tablet	15/10/17-14/11/17	Not relevant
5. Google Analytics	Categories derived by landing page-title		Number of visits	Desktop, Mobile, Tablet	15/10/17-14/11/17; 2017	Kolzchut

Annex 7: “How-Much” aspect – rights realization volume across sources

Tool / Data source	**Metric**	**Employee rights: Most dominant; Least dominant**	**Life event rights: Most dominant; Least dominant**	**Remarks**
1. Buzzilla	Number of conversations	Minimum wage; Pension insurance	Disabilities; Maternity	
2. Ifat’s Panel data sub-set	Number of visits in Kolzchut & NII websites	Minimum wage; Pension Insurance	Disabilities; Unemployment	23% of the panel users used Kolzchut or NII websites
3.1 SimilarWeb – Keywords Analysis	Number of searches	Minimum wage; Pension insurance	Elderly; Disabilities	
3.2 SimilarWeb – Website Analysis	Number of visits in Kolzchut & NII websites	Minimum wage; Advanced training fund	Unemployment; Maternity	
4. Google Trends	Number of searches	Minimum wage; Convalescence pay	Maternity; Unemployment	
5. Google Analytics	Number of visits in Kolzchut website	Minimum wage; Advanced training fund	Maternity; Elderly	

Annex 8: “Where” aspect – reflection of “Where” issues across sources

Data source	Where issues			
	Social media channel use	Websites’ use	Location of use - home vs. work	Device use analysis (desktop vs. tablet vs. mobile)
1. Buzzilla	+ (social networks, forums, blogs, articles, Twitter). Highest use by Facebook	+	-	-
2. Ifat’s Panel data sub-set	-	+ (NII use > Kolzchut use)	+	Desktop only
3. SimilarWeb	+ (social networks only) Highest use by Facebook	+ (NII use > Kolzchut use)		Desktop vs. Mobile, Desktop only in demographic reports. (Mobile use > Desktop use on both NII and Kolzchut websites; Desktop use > Mobile use for life-event rights searches (which referred to NII & Kolzchut websites)
4. Google Trends	-	+ Websites title as search terms (NII use > Kolzchut use)	-	-
5. Google Analytics	-	-	-	+ (Mobile use > Desktop use)

Annex 9: “When” aspect – reflection of “When” issues across sources

Data source	Time range selection options for data extraction	Time granularity (Month / Week / Day / Hour)	Remarks
1. Buzzilla	Last day / week / 3 weeks / month / 3 months / 6 months / Year/ In date / Dates range. Limited to last 2 years	M / W / D	
2. Ifat’s Panel data sub-set	15/10/17-14/11/17	W / D / H	(Highest use on Sundays, lowest use on weekends. During daytime, highest use around 10:00 am.)
3. SimilarWeb	Last 28 days / month / 3/6/12/18/24 months / Months range	M / D (available on some reports) In Keywords Analysis: M	
4. Google Trends	Last hour/ 4 hours / day / 7 / 30 / 90 days / 12 months / 5 years /customized dates range	W / D / H Depends on time range	
5. Google Analytics	Real-time / today / yesterday / last week / month/ 7 days / 30 days / customized dates range. Can choose any start time from the website’s starting day. Comparisons between two time periods are possible.	M / W / D / H	

Annex 10: “Why” aspect – reflection of “Why” issues across sources

Data Source	Units	Advantages	Disadvantages
Buzzilla	Conversations texts (of blogs, forums, social media and news media).	Almost full access to users-made textual data. Qualitative “cyberethnography” process (Hampton, 2017) is possible.	Manual time-consuming process of data cleaning should be conducted. No automatic NLP techniques.
SimilarWeb	Search keywords that lead to the specific analyzed websites (Website analysis tool->Traffic sources-> Search->Keywords).	Keywords that directly referred users to websites, might facilitate focusing on specific points of interest by the users. Processing might be easier as search terms are short.	Short terms and keywords that directly referred users to websites might not shed light on the big picture of users’ intentions.

Annex 11: “Who” aspect - reflection of “Who” issues across sources

Data source	Gender	Age	Region	Income Level	Religious Level	Lang.	Can characterize demographic for people who perform activity regarding a specific right?	Remarks
1. Buzzilla	*	-	-	-	-	-	*	*Manual lexical process (gender and “Mediators” indications might be manually derived from the text).
2. Ifat’s Panel data sub-set	** (survey data)	** (survey data)	** (survey data)	** (survey data)	** (survey data)	+	+	**Desktop users only.
3. SimilarWeb	+	+	+ (Countries)	-	-	-	-	Desktop users only.
4. Google Trends	-	-	+ (Countries and cities)	-	-	-	-	
5. Google Analytics	+	+	+ Cities indicate servers’ locations	-	-	+	+	Segments definitions enable applying complex extractions (e.g. male age 65+).

Society

Samuel Neaman Institute
for National Policy Research

Tel. 972-4-8292329 | Fax. 97-4-8231889
Technion - Israel Institute of Technology
Technion City, Haifa 3200003, Israel
www.neaman.org.il