
NORTH ATLANTIC TREATY
ORGANIZATION



AC/323(IST-141)TP/1079

SCIENCE AND TECHNOLOGY
ORGANIZATION



www.sto.nato.int

STO TECHNICAL REPORT

TR-IST-141

Exploratory Visual Analytics

(Analyse visuelle exploratoire)

This is the Technical Report of the NATO IST-141 Research Task Group
“Exploratory Visual Analytics.”



Published February 2023

Distribution and Availability on Back Cover



NORTH ATLANTIC TREATY
ORGANIZATION



AC/323(IST-141)TP/1079

SCIENCE AND TECHNOLOGY
ORGANIZATION



www.sto.nato.int

STO TECHNICAL REPORT

TR-IST-141

Exploratory Visual Analytics

(Analyse visuelle exploratoire)

This is the Technical Report of the NATO IST-141 Research Task Group
"Exploratory Visual Analytics."

The NATO Science and Technology Organization

Science & Technology (S&T) in the NATO context is defined as the selective and rigorous generation and application of state-of-the-art, validated knowledge for defence and security purposes. S&T activities embrace scientific research, technology development, transition, application and field-testing, experimentation and a range of related scientific activities that include systems engineering, operational research and analysis, synthesis, integration and validation of knowledge derived through the scientific method.

In NATO, S&T is addressed using different business models, namely a collaborative business model where NATO provides a forum where NATO Nations and partner Nations elect to use their national resources to define, conduct and promote cooperative research and information exchange, and secondly an in-house delivery business model where S&T activities are conducted in a NATO dedicated executive body, having its own personnel, capabilities and infrastructure.

The mission of the NATO Science & Technology Organization (STO) is to help position the Nations' and NATO's S&T investments as a strategic enabler of the knowledge and technology advantage for the defence and security posture of NATO Nations and partner Nations, by conducting and promoting S&T activities that augment and leverage the capabilities and programmes of the Alliance, of the NATO Nations and the partner Nations, in support of NATO's objectives, and contributing to NATO's ability to enable and influence security and defence related capability development and threat mitigation in NATO Nations and partner Nations, in accordance with NATO policies.

The total spectrum of this collaborative effort is addressed by six Technical Panels who manage a wide range of scientific research activities, a Group specialising in modelling and simulation, plus a Committee dedicated to supporting the information management needs of the organization.

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS System Analysis and Studies Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These Panels and Group are the power-house of the collaborative model and are made up of national representatives as well as recognised world-class scientists, engineers and information specialists. In addition to providing critical technical oversight, they also provide a communication link to military users and other NATO bodies.

The scientific and technological work is carried out by Technical Teams, created under one or more of these eight bodies, for specific research activities which have a defined duration. These research activities can take a variety of forms, including Task Groups, Workshops, Symposia, Specialists' Meetings, Lecture Series and Technical Courses.

The content of this publication has been reproduced directly from material supplied by STO or the authors.

Published February 2023

Copyright © STO/NATO 2023
All Rights Reserved

ISBN 978-92-837-2396-7

Single copies of this publication or of a part of it may be made for individual use only by those organisations or individuals in NATO Nations defined by the limitation notice printed on the front cover. The approval of the STO Information Management Systems Branch is required for more than one copy to be made or an extract included in another publication. Requests to do so should be sent to the address on the back cover.

Table of Contents

	Page
List of Figures	vi
List of Tables	ix
List of Acronyms	x
Acknowledgements	xiii
IST-141 Membership List	xiv
Executive Summary and Synthèse	ES-1
Chapter 1 – Introduction	1-1
1.1 Background	1-1
1.2 Objectives of the NATO IST-141 RTG	1-1
1.3 Aim of the Report	1-1
1.4 Report Structure	1-2
1.5 References	1-3
Chapter 2 – Human Factors Considerations for Visual Analytics	2-1
2.1 Defining Human Factors	2-1
2.2 Determining What to Present	2-1
2.2.1 User-Centered Design	2-1
2.2.2 Myths About Design	2-3
2.2.3 Users of Visual Analytics Systems	2-4
2.3 Determining How to Present It	2-5
2.3.1 Standards, Guidelines, Heuristics, and Best Practices	2-5
2.3.2 Example: Stereoscopically Perceivable 3D Data Visualizations for Cyber Security	2-6
2.3.3 Virtual Data Explorer	2-6
2.4 Evaluating Design Effectiveness	2-9
2.4.1 Subjective versus Objective Human Performance Measures	2-9
2.4.2 Situation Awareness Measures	2-9
2.4.3 Workload Measures	2-10
2.5 Summary	2-10
2.6 References	2-10
Chapter 3 – Information Visualization and Visual Analytics for the Maritime Domain	3-1
3.1 Introduction	3-1
3.2 Maritime Datasets	3-1

3.3	State of the Art on Maritime Visualization	3-4
3.3.1	Interactive Visualization of Vessel Traffic for Monitoring and Exploratory Analysis	3-5
3.3.2	Visual Analytics for Maritime Pattern Detection	3-7
3.3.3	Three-Dimensional Visualization of Maritime Pattern and Mobility	3-10
3.4	Experiences on Maritime Interactive Visualizations and Visual Analytics	3-12
3.4.1	Maritime Patterns-of-Life Information Service (MPoLIS)	3-12
3.4.2	A Visual Analytics Experience for Naval Command and Control Application: Case Study on the Turkish Straits	3-13
3.4.3	Maritime Cyber Security	3-15
3.5	Conclusion and Discussion	3-17
3.6	References	3-17
Chapter 4 – Exploratory Media Analysis		4-1
4.1	Data and Methods	4-1
4.1.1	Corpora	4-1
4.1.2	Methods	4-1
4.2	Results	4-2
4.2.1	Word Co-Occurrence Networks	4-2
4.2.2	Sentiment-Based Storyline Analysis	4-2
4.3	Conclusions	4-2
4.4	References	4-5
Chapter 5 – Visual Exploration of Simulation Data		5-1
5.1	Introduction	5-1
5.2	Visual Analytics for Simulation Data	5-1
5.2.1	Post Analysis	5-2
5.2.2	In situ Analysis	5-3
5.3	Software Tools Summary	5-4
5.3.1	Time Line Graphs	5-4
5.3.2	The Visualization Toolkit and ParaView	5-6
5.3.3	VisIt	5-8
5.4	Conclusions and Discussion	5-8
5.5	References	5-8
Chapter 6 – Exploring Deep Learning		6-1
6.1	Interactive Visualization and Deep Learning Research	6-1
6.2	Understanding Deep Neural Networks Internal Operations	6-1
6.3	Explaining Deep Learning Results	6-6
6.4	Exploiting the Synergy Between Visual Analytics and Deep Learning	6-9
6.5	Conclusion	6-10
6.6	References	6-11

Chapter 7 – Cyber Situation Awareness	7-1
7.1 Introduction	7-1
7.2 Cyber Situation Awareness	7-1
7.3 Human Machine Interface Design Approaches	7-2
7.4 Symbology	7-4
7.4.1 Symbology, Geo-Spatial and Cyber-Spatial Thinking	7-4
7.4.2 MIL-STD-2525D and NATO APP-6	7-5
7.4.3 Other Cyber Symbology Approaches	7-6
7.5 Conclusions	7-7
7.6 References	7-8
Chapter 8 – Improvised Explosive Device Incidents Analysis with Storytelling Exploratory Visual Analytics	8-1
8.1 Introduction	8-1
8.2 Datasets Descriptions	8-1
8.2.1 NATO Ukraine IED Incidents Data	8-1
8.2.2 Ukraine Census Data	8-1
8.3 Exploratory Visual Analytics Storytelling Tool	8-2
8.3.1 Design Goal and Approach	8-2
8.3.2 Dataset Overview	8-2
8.3.3 Geospatial View	8-2
8.3.4 Incident Type View	8-2
8.3.5 Text Analysis View	8-2
8.4 Storytelling Techniques	8-6
8.4.1 Drill-Down Story	8-7
8.4.2 Martini Glass Structure	8-7
8.4.3 Interactive Slideshow	8-7
8.5 Insights About Improvised Explosive Devices Incidents	8-7
8.6 Conclusion	8-11
8.7 References	8-11
Chapter 9 – HFM-259 Data Exploration	9-1
9.1 Introduction	9-1
9.2 Dataset	9-1
9.3 Visualization Framework	9-2
9.4 Data Preprocessing – Bayesian Network	9-2
9.5 Analytics Dashboards	9-3
9.5.1 Facet Exploration	9-3
9.5.2 Bayesian Network Exploration	9-5
9.6 Conclusion	9-7
9.7 References	9-8
Chapter 10 – Conclusions and Recommendations	10-1
10.1 Conclusions	10-1
10.2 Recommendations	10-2

List of Figures

Figure		Page
Figure 2-1	User-Centered Design	2-2
Figure 2-2	NATO CCDCOE Locked Shields CDX Networks Topology Rendered from NATO CCDCOE Locked Shields 2018 Partner Run Dataset, Overlaid with Activity	2-7
Figure 2-3	Same Constellation as 2-2, but Camera Viewpoint Turned 90 Degrees Clockwise and Moved Behind the “Simulated Internet” Data-Shape; Highlighted (Red) Are Edges Illustrating the Connections of Red Team Activities in Third Blue Team’s Drone Control Nodes	2-7
Figure 2-4	Display of a Blue Team’s Network Topology and Observed Connections During a Time Window	2-8
Figure 2-5	Display of a Blue Team’s Network Layout Where Entities’ Positions on XYZ Axes are Determined by: Z) The Group this Entity Belongs to (a Subnet); Y) Subgroup (a Functional Group in that Subnet: Servers, Networks Devices, Workstations); X) Entity’s Sequential (Arbitrary) Position in that Subgroup (for Example the Last Octet of its IP Address)	2-8
Figure 3-1	Maritime Open Data Available for Research from Zenodo	3-3
Figure 3-2	Example of Data Included in the Dataset	3-4
Figure 3-3	Time Bars for the Analysis of Temporal Variation of Speed in Vessel Trajectories	3-6
Figure 3-4	Visualization Integrating AIS Information (Triangular Icons) and Radar Contacts (Image in Yellow)	3-6
Figure 3-5	Interactive Visualization of Vessel Trajectory for Prediction	3-7
Figure 3-6	Detection of Vessel Meeting Points: For a Given Ship, the Graph Illustrates the List of Visited Ports and the List of Other Ships that Were in the Same Port at the Same Time	3-7
Figure 3-7	Clustering and Discrete Aggregation to Identify Vessel Traffic Lanes and Flows	3-8
Figure 3-8	Analysis of Near Collision Events in the Port of Brest	3-9
Figure 3-9	Spatio-Temporal Visualization for the Analysis of Speed Variations in Vessel Trajectories	3-11
Figure 3-10	Space-Time Cube for Vessel Event Detection: Shift Proximity (Left) and Drifting (Right)	3-11
Figure 3-11	Space-Time Cube for the Visualization of Outliers in Vessel Traffic	3-12
Figure 3-12	MPoLIS Interface, Showing Vessel Traffic for Italian Ports	3-13
Figure 3-13	VATOZ [®] Visualizations	3-14
Figure 3-14	VATOZ [®] Visualizations	3-14

Figure 3-15	VATOZ [®] Visualizations	3-15
Figure 3-16	Web-Based Interface	3-16
Figure 3-17	Detection and Visualization of an Alert	3-16
Figure 4-1	Word Co-Occurrence Network: 1st Stage of the Ukrainian Conflict	4-3
Figure 4-2	Word Co-Occurrence Network: 3rd Stage of the Ukrainian Conflict	4-3
Figure 4-3	Sentiment-Based Narrative Trajectory (with 3 Types of Smoothing: Grey Line – Moving Average, Blue – Loess, Red – Suyzhet Discrete Cosine Transformation): 3rd Stage of the Ukrainian Conflict	4-4
Figure 5-1	Basic Principle of a Plane-Based PCP	5-4
Figure 5-2	Time Line Graphs Example, Showing the Table Data Importer (Top) and the Multi Tabbed Graph View (Bottom)	5-5
Figure 5-3	ParaView Example Showing a Split View, Where Each View Can be Used to Emphasize Different Features in the Data, Both by Shape and Color Schemes	5-7
Figure 5-4	ParaView Example Showing CT Data in a Volume Rendering (Left), an X-Ray Like Slice View (Top Right), a Data Histogram (Bottom Right) and a Transfer Function Editor (Rightmost Column)	5-7
Figure 6-1	A Visual Overview of Interrogative Questions About VA in DL	6-2
Figure 6-2	CNNVis: The Bottom Shows a Screenshot from the Interactive Visualization Software	6-3
Figure 6-3	The Interface of RNNVis	6-4
Figure 6-4	LSTMVis User Interface	6-4
Figure 6-5	ActiVis Integrates Several Coordinated Views to Support Exploration of Complex Deep Neural Network Models, at Both Instance and Subset-Level	6-5
Figure 6-6	Explainable Artificial Intelligence Concept as Presented by DARPA XAI	6-6
Figure 6-7	Left: Husky Classified as Wolf, Right: Explanation of the Model’s Prediction in the “Husky vs Wolf” Task	6-7
Figure 6-8	The Input Image is Correctly Classified as “Rooster”	6-7
Figure 6-9	Explaining Predictions of AI Systems	6-8
Figure 6-10	Joint Classification and Explanation Model Architecture	6-9
Figure 6-11	Visual Explanations Generated by the Model, Containing Image Relevant Sentences with Class Discriminative Attributes	6-9
Figure 6-12	Exploring Clustering Result of VAST 2017, There are Four Outstanding Patterns and Such Patterns Visualized by Heat Map, 3D Map and Relationship Map	6-10
Figure 6-13	Interactive Image Generation	6-10

Figure 7-1	User-Centric Approach	7-3
Figure 7-2	System-Based – Ecological Interface Design	7-3
Figure 7-3	Moving from Geo-Spatial to Cyber-Spatial Visual Representations is a Requirement for Successful Cyber SA	7-5
Figure 7-4	Examples of MIL-STD-2525D Geo-Spatial Symbols (Left) and Cyber-Spatial Symbols Created Independently in the Absence of a Standard (Right)	7-6
Figure 8-1	C-IED Analysis Tool Introduction View	8-3
Figure 8-2	C-IED Analysis Tool Geospatial View	8-4
Figure 8-3	C-IED Analysis Tool Incident View	8-5
Figure 8-4	C-IED Analysis Tool Text Analysis View	8-6
Figure 8-5	Geospatial Summary Slide Featuring a Map of Ukraine IED Incidents in 2014 – 2015	8-8
Figure 8-6	Categorical and Temporal Summary Slides Featuring Statistical Data About Ukraine IED Incidents in 2014 – 2015	8-8
Figure 8-7	Regions Colored According to Level of IED Incidents	8-9
Figure 8-8	Regions Colored According to Lethality of IED Incidents	8-9
Figure 8-9	Interactive Timeline Banner and Filter at the Top of Each View	8-9
Figure 8-10	Sankey Diagram Showing Incidents by Type and Outcome	8-10
Figure 8-11	Sankey Diagram Showing Casualties by Incident Type	8-10
Figure 9-1	Dashboard with Widgets for Facet Exploration with Tag-Clouds in Circle Pack Layout	9-3
Figure 9-2	Dashboard with Widgets for Facet Exploration with Tag-Clouds in Horizontal Bar Layout	9-4
Figure 9-3	Facet Exploration by Elements of Taxonomy – Details	9-4
Figure 9-4	Facet Exploration by Elements of Taxonomy – Details	9-5
Figure 9-5	Bayesian Network for All Occurrences of All Values (Left) and Aggregated to Category-Elements (Middle) Bar Chart as Legend for Color Code of Category	9-6
Figure 9-6	Selecting Nodes in Network Diagram to Show Distribution of Co-Occurrences in (Normalized) Sankey	9-6
Figure 9-7	Normalized Sankeys	9-7

List of Tables

Table		Page
Table 2-1	Common Human Factors Design Standard Topics	2-5
Table 2-2	Measures of SA	2-9

List of Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
ACM	Association for Computing Machinery
AI	Artificial Intelligence
AIS	Automatic Identification System
APP	Allied Procedural Publication
AR	Augmented Reality
ARL	U.S. Army Research Laboratory
ASCII	American Standard Code for Information Interchange
C2	Command and Control
C5ISR	Command, Control, Computers, Communications, Cyber, Intelligence, Surveillance, Reconnaissance
CBRNE	Chemical, Biological, Radiological, Nuclear, Explosive
CC	Creative Commons
CCDC	Combat Capabilities Development Command
CCDCOE	Cooperative Cyber Defence Centre of Excellence
CDX	Cyber Defence Exercise
CFD	Computational Fluid Dynamics
C-IED	Counter Improvised Explosive Device
CMRE	Centre for Maritime Research and Experimentation
CNN	Convolutional Neural Network
COE	Centre of Excellence
Cranfield-SAS	Cranfield Situation Awareness Scale
CSO	Collaboration Support Office
CSSP	Cyber Security Service Provider
CT	Computed Tomography
datAcron	Big Data Analytics for Time Critical Mobility Forecasting
DDOS	Distributed Denial-of-Service
DGMs	Deep Generative Models
DKOE	Decision Knowledge Operational Effectiveness
DNS	Domain Name System
DoD	Department of Defense
DQN	Deep Q-Network
DRDC	Defence Research & Development Canada
DS	Decision Support
EID	Ecological Interface Design
ESS	Earth System Science
GANs	Generative Adversarial Networks
GUI	Graphical User Interface
HFM	Human Factors and Medicine
HPC	High Performance Computing
IA	Information Analysis
IED	Improvised Explosive Device

IEEE	Institute of Electrical and Electronics Engineers
IMO	International Maritime Organization
IST	Information Systems Technology
ITU-R	International Telecommunication Union – Recommendation
IVIS	Interactive Visualizations
KIA	Killed in Action
LIME	Local Interpretable Model-agnostic Explanations
LRP	Layer-wise Relevance Propagation
LSTM	Long Short-Term Memory
MARS	Mission Awareness Rating Scale
MIL-STD	Military Standard
MLP	Multi-Layer Perceptron
MOD	Ministry of Defence
MPoLIS	Maritime Patterns-of-Life Information Service
MR	Mixed Reality
MSA	Maritime Situational Awareness
NATO	North Atlantic Treaty Organization
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
NMEA	National Marine Electronic Association
NMSG	NATO Modelling and Simulation Group
OSeaM	Open Sea Map
OSM	Open Street Map
PCP	Parallel Coordinates Plot
QUASA	Quantitative Analysis of Situation Awareness
RNN	Recurrent Neural Network
RTG	Research Task Group
RTO	Research and Technology Organization
SA	Situation Awareness
SABARS	Situation Awareness Behavioral Rating Scale
SAGAT	Situation Awareness Global Assessment Technique
S-AIS	Satellite AIS
SAR	Search and Rescue
SART	Situation Awareness Rating Technique
SAS	System Analysis and Studies
SA-SWORD	Situation Awareness Subjective Workload Dominance
SAVANT	Situation Awareness Verification and Analysis Tool
SME	Subject Matter Expert
SN	Social Networks
SNA	Social Network Analysis
SOM	Self-Organizing Map
SPAM	Situation Present Assessment Method
STANAG	Standardization Agreement
SWAT	Subjective Workload Assessment Technique

T-AIS	Terrestrial AIS
TCT	Trajectory Contingency Tables
TLG	Timeline Graphs
TLX	Task Load Index
t-SNE	t-distributed Stochastic Neighbor Embedding
UCD	User-Centered Design
UI	User Interface
UK	United Kingdom
US	United States
UX	User Experience
VA	Visual Analytics
VAST	Visual Analytics Science and Technology
VDE	Virtual Data Explorer
Vids	Visual Intrusion Detection System
VizSec	Visualization for Cyber Security
VMS	Vessel Monitoring System
VR	Virtual Reality
VRDAE	Virtual Reality Data Analysis Environment
VTK	Visualization Toolkit
WIA	Wounded in Action
xR	Extended Reality

Acknowledgements

We would like to acknowledge the invaluable support and guidance from our mentor and IST Panel Chair Dr. Michael Wunder and IST Panel Vice Chair Col. Dr. Nikolai Stoianov, UK IST panel representatives Mr. Simon Baker and Prof. Bob Madahar, also the invaluable support from Col. (rtd) Philippe Soète, Ms. Aysegul Apadin, Ms. Agata Fernandes Swiatkiewicz, Ms. Armelle Dutruc and Mr. Bernard Garcin from CSO.

IST-141 Membership List

CHAIR

Dr. Margaret VARGA*
University of Oxford
UNITED KINGDOM
Email: margaret.varga@zoo.ox.ac.uk

MEMBERS

Mr. Ethem ARKIN*
Aselsan
TURKEY
Email: earkin@aselsan.com.tr

Dr. Petter BIVALL*
Swedish Defence Research Agency (FOI)
SWEDEN
Email: petter.bivall@foi.se

Dr. Elena CAMOSSO*
CMRE
CMRE – Centre for Maritime Research and
Experimentation
Email: lena.camossi@cmre.nato.int

Mr. Kaur KULLMAN*
Estonian Ministry of Defense
ESTONIA
Email: kaur@ieee.org

Dr. Tomas KRILAVIČIUS*
Vytautas Magnus University (Vytauto Didžiojo
Universitetas)
LITHUANIA
Email: tomas.krilavicius@krilas.lt

Mrs. Valérie LAVIGNE*
Defence Research and Development Canada –
Valcartier
CANADA
Email: valerie.lavigne@drdc-rddc.gc.ca

Dr. Kristen LIGGETT*
Air Force Research Laboratory, Airman Systems
Directorate
UNITED STATES
Email: kristen.liggett@us.af.mil

Dr. Virginijus MARCINKEVIČIUS*
Vilnius University
LITHUANIA
Email: virginijus.marcinkevicius@mii.vu.lt

Mr. Cyril RAY*
Ecole navale (French Naval Academy)
FRANCE
Email: cyril.ray@ecole-navale.fr

Ms. Susan TRAEBER-BURDIN*
Fraunhofer FKIE
GERMANY
Email: susan.traeber-burdin@fkie.fraunhofer.de

Dr. Carsten WINKELHOLZ*
Fraunhofer FKIE
GERMANY
Email: carsten.winkelholz@fkie.fraunhofer.de

* Contributing or Supporting Author

ADDITIONAL CONTRIBUTORS

Mr. Nikhil ACHARYA*
Fraunhofer FKIE
GERMANY
Email: nikhil.acharya@fkie.fraunhofer.de

PANEL/GROUP MENTOR

Dr. Michael WUNDER
Fraunhofer-FKIE
GERMANY
Email: michael.wunder@fkie.fraunhofer.de

* Contributing or Supporting Author



Exploratory Visual Analytics

(STO-TR-IST-141)

Executive Summary

Information superiority is one of the primary enablers for military dominance; the exploitation of all relevant information from multiple sources is a key factor for NATO's information superiority. Visualization and visual analytics research are essential to address the needs of the 2015 NATO targets of emphasis in Information Analysis (IA) and Decision Support (DS): IA&DS-1 on Decision Support and IA&DS-2 on Big Data and Long Data Processing and Analysis.

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. The IST-141/RTG-66 group investigated, researched, and fostered collaborations in visual analytics and visualization – facilitating knowledge extraction and data analysis for timely situation awareness and effective decision making. The objectives of IST-141 were thus to research, develop and apply exploratory visual analytics techniques:

- 1) To exploit and make sense of large and complex data, i.e., Big data;
- 2) To help make tacit knowledge explicit;
- 3) To provide acute situation awareness; and
- 4) To support informed decision making across a wide range of defence domains including cyber, maritime, genomics, and social media domains, as well post analysis and in situ visualization for simulation data.

In addition, IST-141 facilitated the interest in, as well as the uptake and exploitation of, visual analytics technologies within and beyond NATO through:

- 1) Organizing and presenting at one NATO Specialists' Meeting (IST-HFM-154: Cyber Symbolology) and one NATO inter-panel and inter-group workshop (IST-178: Big Data Challenges: Situation Awareness and Decision Support).
- 2) Lecturing on two NATO Lecture Series (IST-143 and IST-170) during 2016 – 2019 in eight different countries.
- 3) Contributing to, and participating in, NATO CSO activities organized by others and initiating/participating in joint activities.
- 4) Presenting at many prestigious international conferences, workshops, and seminars such as IEEE VIS.
- 5) IST-141 members also work widely outside the group in collaborations with IST-108, IST-129, IST-177, IST-ET-094, IST-ET-099, IST-SAS-102, HFM-259, HFM-294, SAS-124, SAS-117 and SAS-139.

The group generated 32 publications.

Analyse visuelle exploratoire

(STO-TR-IST-141)

Synthèse

La supériorité en matière d'information est l'un des principaux outils de prédominance militaire ; l'exploitation de toutes les informations pertinentes de multiples sources est un facteur clé pour l'OTAN dans ce domaine. Les recherches sur la visualisation et l'analyse visuelle sont essentielles pour atteindre les objectifs prioritaires de l'OTAN en 2015 dans le domaine de l'analyse de l'information (AI) et de l'aide à la décision (DS) (IA&DS-1 sur l'aide à la décision et IA&DS-2 sur les données massives et le traitement et l'analyse de données longues).

L'analyse visuelle est la science du raisonnement analytique facilitée par des interfaces visuelles interactives. Le groupe IST-141/RTG-66 a mené des études et des recherches et favorisé des collaborations en analyse visuelle et visualisation, en facilitant l'extraction des connaissances et l'analyse des données afin d'établir une connaissance de la situation opportune permettant une prise de décision efficace. Les objectifs de l'IST-141 étaient donc d'étudier, mettre au point et appliquer des techniques d'analyse visuelle :

- 1) Pour exploiter et donner un sens aux données vastes et complexes, autrement dit, aux données massives ;
- 2) Pour faciliter l'explicitation de l'implicite ;
- 3) Pour fournir une connaissance de la situation précise ; et
- 4) Pour favoriser une prise de décision éclairée dans une large palette de domaines de la défense, y compris le cyberdomaine, le domaine maritime, la génomique et les médias sociaux, ainsi pour la post-analyse et la visualisation in situ des données de simulation.

De plus, l'IST-141 a renforcé l'intérêt pour et facilité l'essor et l'exploitation des technologies d'analyse visuelle au sein de l'OTAN et en dehors, par :

- 1) L'organisation d'une réunion des spécialistes (IST-HFM-154, « Cybersymbologie ») et d'un séminaire intercommission et intergroupe (IST-178, « Défis des données massives : connaissance de la situation et aide à la décision ») et la présentation d'exposés à ces occasions ;
- 2) L'intervention dans deux séries de conférences OTAN (IST-143 et IST-170) entre 2016 et 2019 dans huit pays ;
- 3) La contribution et la participation aux activités du CSO de l'OTAN organisées par d'autres et le lancement ou la participation à des activités conjointes ;
- 4) La présentation d'exposés dans un grand nombre de conférences et séminaires internationaux prestigieux, tels que VIS de l'IEEE ;
- 5) Les membres de l'IST-141 travaillent par ailleurs largement en dehors du groupe, au sein de collaborations avec l'IST-108, l'IST-129, l'IST-177, l'IST-ET-094, l'IST-ET-099, l'IST-SAS-102, le HFM-259, le HFM-294, le SAS-124, le SAS-117 et le SAS-139.

Le groupe a produit 32 publications.

Chapter 1 – INTRODUCTION

Margaret Varga

University of Oxford / Seetru Ltd.
UNITED KINGDOM

Petter Bivall

Swedish Defence Research Agency
SWEDEN

Kristen K. Liggett

Air Force Research Laboratory
UNITED STATES

Valérie Lavigne

Defence Research and Development Canada
CANADA

1.1 BACKGROUND

Information superiority is one of the key elements for military dominance; the exploitation of all relevant information from multiple sources is a crucial factor for NATO's information superiority. Visualization and Visual Analytics research are essential to address the needs of the 2015 NATO targets of emphasis in Information Analysis (IA) and Decision Support (DS): IA&DS-1 on Decision Support and IA&DS-2 on Big Data and Long Data Processing and Analysis.

Visual Analytics (VA) is the science of analytical reasoning facilitated by interactive visual interfaces [1]. There are three main components of VA, namely, interactive visualization, analytical reasoning, and computational analysis [2]. In the context of VA considered by the Group:

- Visualization is concerned with the use of interactive visual representations of data to amplify cognition [3], while;
- Analytical reasoning and computation analysis work to support data exploration, analysis and understanding.

1.2 OBJECTIVES OF THE NATO IST-141 RTG

The NATO IST-141/RTG66 Research Task Group (RTG) Exploratory Visual Analytics investigated, researched and fostered collaborations in knowledge extraction/discovery and data analysis for timely situation awareness to support effective decision making. The Group explored how visualization conveys information effectively: leveraging human perception and enhancing human cognition, i.e., bringing together visualization and the user's mental model (see Chapter 2 and [4]). The objectives of IST-141 were thus to research, develop and apply exploratory visual analytics techniques:

- 1) To exploit and make sense of large and complex data sets, i.e., Big data;
- 2) To help make tacit knowledge explicit;
- 3) To provide acute situation awareness, and
- 4) To support informed decision making in a wide range of different defence domains, such as maritime, social media, genomics, and cyber domains as well as post analysis and in situ visualization for simulation data.

1.3 AIM OF THE REPORT

The aim of this IST-141 Research Task Group technical report is to discuss the work conducted by the Group to research, develop and apply exploratory visual analytics to data sets relating to:

INTRODUCTION

- Improvised explosive devices resources (NATO C-IED COE);
- Social media; and
- Cyber and maritime and intelligence operations.

The work demonstrates the effectiveness of Exploratory VA in detecting, monitoring, analyzing, and understanding large and complex datasets, i.e., big data, for situation awareness and decision support.

This report will also discuss the work of the Group on:

- 1) Research and development of visualization and visual analytics technologies.
- 2) Raising awareness of the Group's work:
 - By presenting papers at prestigious international conferences, e.g., IEEE VIS; and
 - Contributing to two NATO lecture series (IST-143 and IST-170).
- 3) Facilitating the exploitation and application of visual analytics and visualization technologies in NATO defence and security domains, and beyond.
- 4) Broadening the horizon of the understanding and exploration of visualization and visual analytics.
- 5) Leveraging the generation of new ideas.
- 6) Developing NATO inter-Panel/inter-Group collaborations through:
 - Organization (and presentation of the Group's work therein) of one joint Panel NATO Specialists' meeting (IST-HFM-154: Cyber Symbology) and one NATO inter-Panel / inter-Group Workshop (IST-178: Big Data Challenges – Situation Awareness and Decision Support);
 - Participating in NATO activities organized by others; and
 - Organizing joint meetings with numerous RTGs from different Panels.

1.4 REPORT STRUCTURE

The chapters in the report summarize work conducted during the course of this RTG.

Chapter 2 discusses human factors considerations for visual analytics. It begins by defining human factors and describes the human factors / user-centered design process. It discusses some common myths about the design process for designers to be aware of and avoid. Users of visual analytics systems are many and diverse, so knowing the user for any project is of utmost importance to ensure that the output product is both useful and usable. References to standards, guidelines, heuristics, and best practices for how to optimally display information are provided. Included in the chapter is a discussion and figures depicting the advantages of using stereoscopic three-dimensional visualizations for particular data sets. Finally, there is a section on how to evaluate the usefulness and usability of visualizations. Included are resources for situation awareness and workload metrics.

Chapter 3 discusses Information visualization and visual analytics in the maritime domain.

Chapter 4 and Chapter 5 are concerned with social media data and simulation data.

Chapter 6 discusses the interactions between visual analytics and deep learning.

Chapter 7 discusses the Group's work in cyber situation awareness and cyber symbology.

Chapter 8 and Chapter 9 explore the application of visual analytics and visualization to NATO data such as:

- IED (NATO C-IED COE); and
- NATO HFM-259 data.

These two chapters discuss the development of, and resulting design principles for, web-based access to these datasets for a wide range of users from the general public to researchers and policy makers, i.e., people from different backgrounds and with varying levels of expertise and knowledge. The analysis of IED data adopted an interactive storytelling approach to engage the general public, and the visual analytics / visualizations of the HFM-259 data are also suitable for public engagement.

Chapter 10 draws conclusions and makes recommendations.

1.5 REFERENCES

- [1] Thomas, J.J. and Cook, K.A. (Eds.) (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center.
- [2] Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J. and Melançon, G. (2008). *Visual Analytics: Definition, Process, and Challenges*. 10.1007/978-3-540-70956-5_7.
- [3] Card, S., Mackinlay, J. and Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers.
- [4] Tory, M. and Moller, T. (2004). *Human Factors in Visualization Research*. In *IEEE Transactions on Visualization and Computer Graphics*, 10(1), pp. 72-84, Jan – Feb 2004.



Chapter 2 – HUMAN FACTORS CONSIDERATIONS FOR VISUAL ANALYTICS

Kristen K. Liggett

Air Force Research Laboratory
UNITED STATES

Kaur Kullman

Cognitive Data OÜ
ESTONIA

2.1 DEFINING HUMAN FACTORS

According to the International Ergonomics Association, “Ergonomics (or human factors) is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and other methods to design in order to optimize human well-being and overall system performance” [1]. California State University Long Beach, Department of Psychology, defines human factors as “...a scientific discipline which examines human behavior and capabilities in order to find the best ways to design products, equipment and systems for maximum safe, effective, satisfying use by humans” [2]. While there are many definitions of human factors in the literature, these two are highlighted because they focus on the consideration of human capabilities in the design of things with which humans interact to maximize *effective performance* and *user satisfaction*. One of the first books published on human factors was an engineering textbook entitled “Human Factors in Engineering and Design” [3]. It was first published in 1957 and is now in its 7th edition. Seminal work in human factors was organized into The Handbook of Human Factors [4]. The book entitled “The Psychology of Everyday Things” (more recently published as “The Design of Everyday Things” [5]) was written in 1988 to introduce these concepts to a wider, non-technical, audience. Although the human factors discipline has been around for quite some time, it is occasionally “rediscovered” and, at various times, has been rebranded as user-centered design, user-driven development, User Experience (UX) design, human-centered design, human engineering, cognitive systems engineering, and most recently, design thinking.

2.2 DETERMINING WHAT TO PRESENT

2.2.1 User-Centered Design

User-Centered Design (UCD) is the cornerstone of the human factors discipline and has been used effectively in many domains, including designs of work support systems for aircraft cockpits, unmanned aerial vehicle control stations, automotive dashboards, nuclear power plant control rooms, space vehicle controls, hospital medical record systems, and many more safety-critical applications.

At the most basic level, UCD is a design approach that definitively places the user at the center of all design activities (Figure 2-1). According to William Hudson of the Interaction Design Foundation [6], UCD is “an iterative design process in which designers focus on the users and their needs in each phase of the design process. UCD calls for involving users throughout the design process via a variety of research and design techniques so as to create highly usable and accessible products for them.” The process starts with a deep understanding of the user and their work. This analysis informs the designer of what to present to the user and *drives* the design process. The analysis phase of the design process also provides designers an opportunity to establish a relationship between the design team and the end-users of the product being designed. These users will supply the design team with valuable information about the work domain, including the overall goal of the work, tasks necessary to accomplish it, the objectives, order, and dependencies of those tasks, information requirements, etc. The design team can use techniques such as interviews, observations, task analysis, and workflow diagramming to gather information for further

analysis. One such technique, goal-directed task analysis, is a procedure of interviews and knowledge elicitation methods that seeks all information related to goals of operators, and the information needed to achieve them in a technologically agnostic manner [7]. Information collection and subsequent analysis help the design team understand the tasks and stakeholders, the information, and decisions to be supported, current sources of information, gaps in current processes, and information gaps. Ultimately, the analysis phase will allow members of the design team to determine how to best support the work processes of end-users in a way that facilitates their cognitive and perceptual needs.

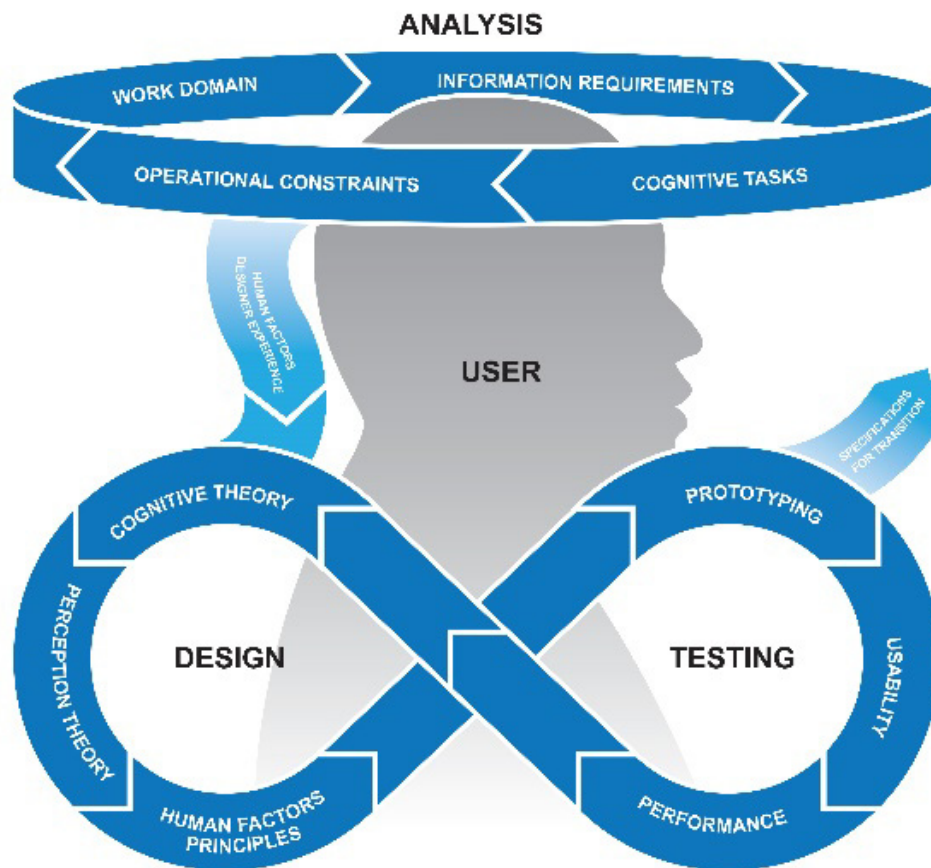


Figure 2-1: User-Centered Design.

Next, and most challenging, is the conversion of information gathered during the analysis phase into initial design concepts (represented by the arrow from the Analysis to Design and Testing Phase in Figure 2-1). This step represents the integration of information gathered in the analysis phase with foundational empirical knowledge of human perception (vision theory, color theory, etc.) and cognition (encoding theory, memory theory, information processing theory, multiple resource theory, attention theory, etc.) guided by human factors design principles that represent decades of research in determining the most effective ways to display information for different uses. Unfortunately, this essential step often lacks a proper foundation, and developers proceed directly to the generation of visualizations without a complete understanding of the work or of how to apply psychological research to best support the end-users’ cognitive capabilities, task goals, or workflow. Too often, complex designs are created under misguided notions, such as that providing end-users vast amounts of information will allow them to find what they need when they need it, that training or documentation will allow users to bridge gaps, or that users’ knowledge and/or expertise about the work itself will be sufficient to allow them to tailor the work aids, interfaces, and/or visualizations in

a way that will support them in their work. Historically, this traditional approach has resulted in systems that fail to provide the expected performance improvements or in systems that are not used at all because they *add* to rather than *support* work.

Once the information from the analysis phase is converted into an initial design, the design team can begin developing the visualization/tool interfaces as well as the tool functionality. The design team will determine sources of required data to support the visualizations and consider methods of effectively accessing the data. The design team will also determine appropriate task allocation to include how automation might be employed. Interaction with end-users is critical for refining initial designs during the design and testing phases. If the previous steps have been done well, designs will require fewer iterations. During this phase, products are tested, refined, and tested again as needed to ensure maximum utility and usability. Latter evaluation involves user-in-the-loop testing with operationally representative scenarios consisting of representative user tasks that have been developed from information gained in the analysis phase.

When UCD is employed by developers with design experience who understand human capabilities and limitations and basic human factors design principles, the resulting designs are effective and intuitive and have a high degree of user acceptance. These designs support workflow without adding unwanted workload.

2.2.2 Myths About Design

Often, tools are built based on factors other than user needs, such as when developers try to find new ways to make use of existing technology or to make use of a new and innovative technology. Other times, user needs drive design, but the designs are not based on a thorough evaluation of the users' processes and goals. While these types of development are not useless, they do not generally lead to the most effective tools and often lead to tools that get shelved because they either don't fit into the users' work processes or because they provide a service that the user doesn't really require. Three common "myths" about how to provide capability are often employed by developers as excuses for circumventing the user-centered design process:

- **Design Myth #1:** Just ask the users what they want and give it to them.
- **Design Myth #2:** Give users access to everything they could possibly need, and they will find what they need when they need it.
- **Design Myth #3:** Allow users to design their own visualizations and interfaces by making everything customizable and letting them choose how they want things to look and behave.

There is solid empirical evidence that indicates that these "myths" are not only inaccurate but can even be dangerous if employed in the design of critical systems.

Design Myth #1 – Give the Users What They Say They Want

While it may be true that users know what they want, that knowledge is often limited by experience (constraints they currently work within) and the lack of knowledge of what is possible. It has been said that if Henry Ford had asked people what they wanted, they would have said 'faster horses' [8]. While there is some debate as to whether this is an actual quote from Ford, it illustrates the fact that getting user input involves more than just asking people what they want. The more important information for designers to obtain when the users are excited about telling them what they want (and they will be!) is the understanding of *why* they are asking for specific things. It is up to designers to probe users for goals, constraints, and motivations. Yes, faster horses may be the users' interpretation of a requirement, but the underlying reason for this request is so they can get from point A to point B faster. Analysis activities as described above supply the design team with a valuable knowledge for design. Using this knowledge, designers can provide tools that allow users to perform work in the best possible way using current technology and sometimes provide requirements that expand current technology rather than tools that just allow them to marginally improve upon current work processes that have been defined and limited by the constraints of past tools and technologies (i.e., faster horses).

Design Myth #2 – Give Users Access to Everything and They Will Find What They Need

This myth stems from the frequent requirement for users to go to a variety of sources to obtain needed information. Designing a system that collects and stores all of this information is an incremental improvement, but that capability alone certainly does not provide an optimal solution. Many dashboards are designed using this myth, and they are frequently not effective for particular users due to the large amount of time and cognitive effort required to sort through and filter information. Often these complex designs are created under the misguided notion that training or documentation will allow users to bridge the gap between the excessive amounts of information and their ability to process excessive amounts of information. As previously mentioned, the most challenging step in UCD is the conversion of domain knowledge obtained during the analysis phase into initial design concepts. This step requires the integration of information gathered in the analysis phase with foundational empirical knowledge of human perception (vision theory, color theory, etc.) and cognition (encoding theory, memory theory, information processing theory, multiple resource theory, attention theory, etc.) guided by human factors design principles that represent decades of empirical research to determine the most effective ways to display information for different uses. This foundational knowledge and experience in applying it to design prevents the creation of complex designs that overtax users' perception and cognition. Design teams must also determine sources of required data to support the interfaces and visualizations and consider methods of effectively accessing those data. Through these activities, a design team can provide an effective tool that will guide users to information they need when they need it, and training and documentation needs will be minimal.

Design Myth #3 – Make Everything Customizable and Let the User Design

One problem with this notion that users typically do not have the time or ability to customize their interfaces. More importantly, multiple studies have been conducted that illustrate that there is often a disconnect between configurations that people say they like best and those with which they perform most effectively [9]. While there are some features of visualizations and interfaces that can and even should remain flexible, the choice of many features is optimal only when their designs and configuration are primarily dependent on workflow needs and goals and fitted to human perceptual and cognitive needs. As indicated above, the conversion of 'work needs' to 'work aids' requires both a deep understanding of the work and a foundational knowledge of human perceptual and cognitive processes along with knowledge of human factors design principles. Users cannot be expected to have this foundational knowledge and should not be made responsible for design. On the other hand, users do have a deeper understanding of the work than can be obtained by the designer, which is why interaction with end-users is critical for refining initial designs. User involvement has the added advantage of ensuring user buy-in. Users often end up preferring things that are best for them when they feel they have been allowed an adequate amount of input into the design process.

2.2.3 Users of Visual Analytics Systems

There are many different types of users of visual analytics systems – but there is **always** a user. The point of visual analytics is to leverage the capabilities of the human visual system to understand complex data. So, whether the users are skilled analysts looking for new insights into complex data sets or the general public trying to understand COVID-19 pandemic information, there are goals, objectives, and tasks unique to each user group. The user groups for visual analytics systems need to be identified and analyzed, and results must be considered when deciding the presentation and interaction details of a visual analytics system. For instance, expert analysts may want to understand information about optimization constraints on algorithms used to process data sets (how do results change if the optimization parameters are weighted differently), while non-experts may want to understand relative comparisons based on personal interests (how does the number of cases/deaths in my state compared to the number in my family's states?). In both of these situations, the presentation of information and how each user group will interact with the information will be very different. If a system will be used by multiple types of users, having a mechanism to tailor the presentation and interaction prior to use based on some initial questionnaire information could allow for pre-use tailoring of the

system. While it is certainly more work to tailor a system to accommodate more than one user group or tailor a system at all, the benefits in terms of portraying useful information and providing a means to explore the data can increase dramatically. Designing a visual analytics system independent of user considerations will lead to unsatisfactory results in terms of both usefulness and usability. One size fits all none.

2.3 DETERMINING HOW TO PRESENT IT

2.3.1 Standards, Guidelines, Heuristics, and Best Practices

Once user analyses have been completed and the designer has identified requirements for what to present, focus now shifts to how to present it. This is an equally important design challenge. There are many resources that designers can reference for guidance in this area. Also, past design experience (examples of things that have worked successfully in the past for a similar requirements) should be considered. Reference documents relevant to human factors design typically come in the form of standards. When there are not established standards, guidelines, best practices, and heuristics suffice. Guidelines are often a useful set of Dos and Don'ts. Best Practices are techniques that have shown to be consistently better than other techniques. Heuristics are general rules of thumb. Jakob Nielsen has generated a useful set of 10 heuristics for user interface design [10].

As mentioned in Section 2.1, the Handbook of Human Factors [4] is a great starting place for guidance and examples of how to display information. There are also numerous design standards that have been produced to guide the designer. The United States (US) Department of Defense MIL-STD-1472 is a military design standard on human engineering [11]. The Ministry of Defence (MOD) Standard 00-250 [12] provides design guidance for the MOD defence acquisition contracts. The International Organization for Standardization established a standard on human-centered design for interactive systems [13] that focuses on ways in which both hardware and software components of interactive systems can enhance human-system interaction. These are just a few of the standards relative to human factors design in existence today. Of particular interest to visual analytics system designers may be the National Institute of Standards and Technology (NIST) Human Engineering Design Criteria Standards [14]. This standard was developed for the Department of Homeland Security and specifically considers use of products by diverse users' groups (civil servants, public health officials, travelers, first responders, and the general public). Visual analytics systems have diverse user groups as well, as pointed out in Section 2.2.3, Table 2-1 shows common topics in many of these standards.

Table 2-1: Common Human Factors Design Standard Topics.

Accessibility	Alarms/Warnings/ Cautions	Anthropometry and Biomechanics	Audio Displays	Communications
Controls	Controls and Display Integration	Dialogue Principles	Environmental Factors	Error Management
Feedback	Forms and Data Entry	Hazards and Safety	Help/Instructions/ Tutorials/Training	Information Coding
Input Devices	Labels	Physical Accommodation	Selection Methods	Signs, Symbols, and Markings
Software Elements	System Status	Use of Automation	Visual Displays	Workstation/Work space Layout

Each Topic has sub-topics. For instance, Visual Displays is commonly broken down into display content and display hardware. This topic was selected for this example because, in visual analytics, 3D representation of data can be essential to understanding the data.

2.3.2 Example: Stereoscopically Perceivable 3D Data Visualizations for Cyber Security

Customized, stereoscopically perceivable 3D visualizations, aligned with cybersecurity analysts' internalized representations of their data, may enhance their capability to understand the state of their networked systems in ways that flat displays with either text, 2D or perspective 3D visualizations cannot afford. For these visualizations to be useful and usable, those need to be aligned to analysts' internalized understanding (mental model) of their data. Section 2.2 described methods for extracting analysts' implicit and explicit understanding of the data that they work with, to create useful, interactive, and importantly for this section, stereoscopically perceivable visualizations that would assist them with their tasks.

Although there have been quite a few recent attempts to utilize Augmented Reality (AR), Mixed Reality (MR) and Virtual Reality (VR) headsets for data visualization, those are usually geared towards showing users the usual (flat) 2D visualizations in Extended Reality (xR) environments or are using relational graphs to show clusters and their relations in 3D. These visualizations can be useful for an initial familiarization phase with a dataset, when a Subject Matter Expert (SME) is trying to learn the functional topology and common behaviors of that dataset but are not particularly useful for interactive data exploration after the user has already gained initial understanding of the topology, relations of its entities and their groups (of groups [of groups (of groups)]) and their expected behavior. Following is a brief overview of a tool, Virtual Data Explorer, which enables the creation of stereoscopically perceivable 3D visualizations and their exploration in mixed or virtual reality environment.

2.3.3 Virtual Data Explorer

For a data visualization that is composed of data-shapes (representations of groups (of groups) of entities in predetermined locations) or their constellations to be useful, the SME must be able to readily map familiar data into a data-shape and choose visual encoding for its attributes so that the resulting visualization will enhance their understanding of that data. Only once a SME is intimate with the composition of the visualization and its relation to the underlying dataset or source, can this SME use that visualization to extract information from the underlying data. To explore the usability of such visualizations, Virtual Data Explorer (VDE) software was created. VDE, which may be employed for visualizing cybersecurity specific datasets, is described in more detail in Refs. [15], [16], [17]. Figure 2-2 through Figure 2-5 show how VDE displays cyber data from the NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) Locked Shields Cyber Defence Exercise (CDX). These figures depict a) Functional topology of entities present in computer networks (22 blue teams, red team's networks, and the game network infrastructure, etc.) that is overlaid with b) Network traffic visualization, where edges (green lines) represent sessions that were observed between entities during a time window, with each edge's opacity referring to the session count between two entities. For detailed explanation of VDE and exercise, see Refs. [15] and [16].

Creation of VDE was partly supported by the Army Research Laboratory under Cooperative Agreement Number W911NF-17-2-0083 and in conjunction with the CCDC Command, Control, Computers, communications, Cyber, Intelligence, Surveillance, and Reconnaissance (C5ISR) Center.

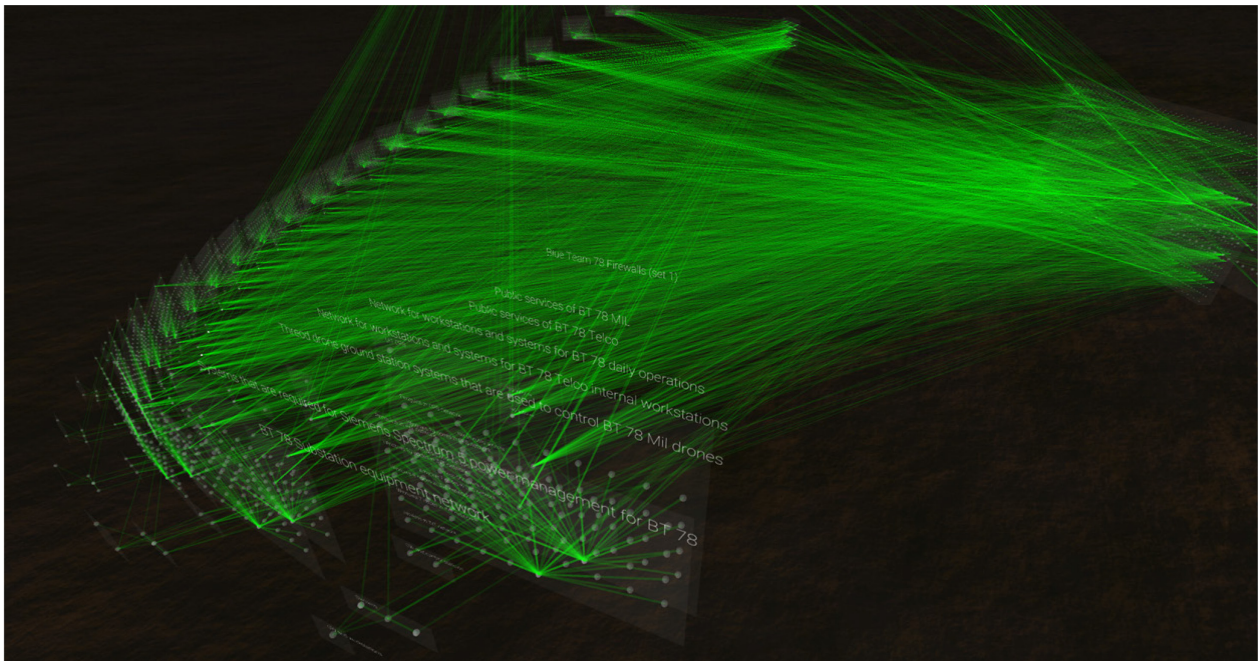


Figure 2-2: NATO CCDCOE Locked Shields CDX Networks Topology Rendered from NATO CCDCOE Locked Shields 2018 Partner Run Dataset, Overlaid with Activity. (Source: Ref. [17].)

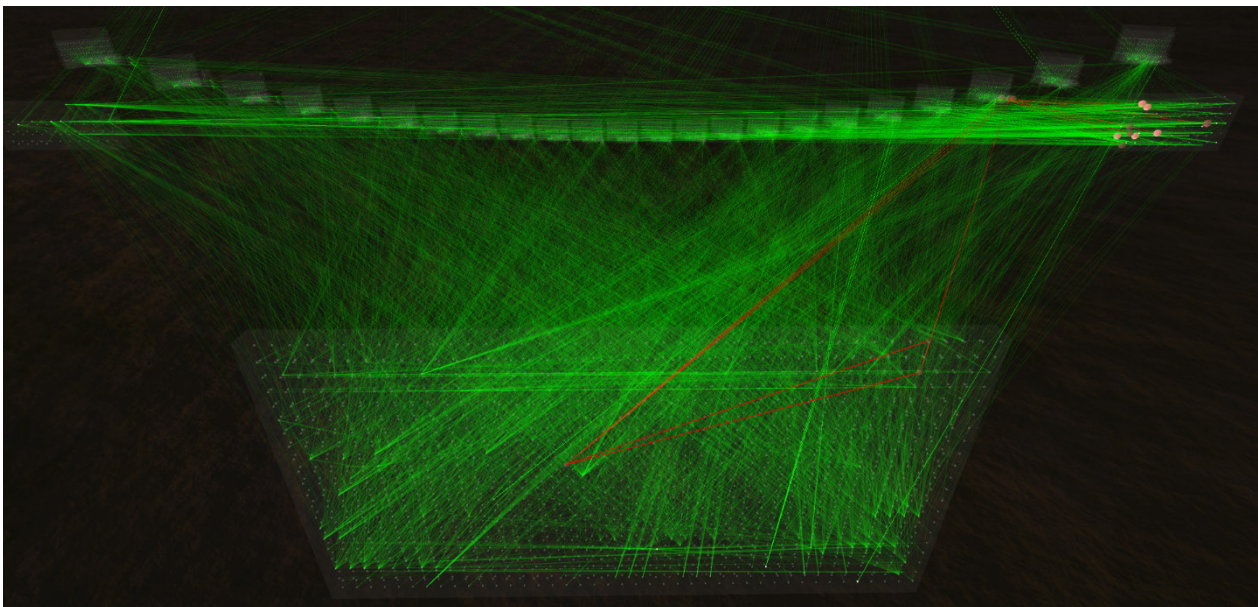


Figure 2-3: Same Constellation as 2-2, but Camera Viewpoint Turned 90 Degrees Clockwise and Moved Behind the “Simulated Internet” Data-Shape; Highlighted (Red) Are Edges Illustrating the Connections of Red Team Activities in Third Blue Team’s Drone Control Nodes. (Source: Ref. [17].)



Figure 2-4: Display of a Blue Team’s Network Topology and Observed Connections During a Time Window.

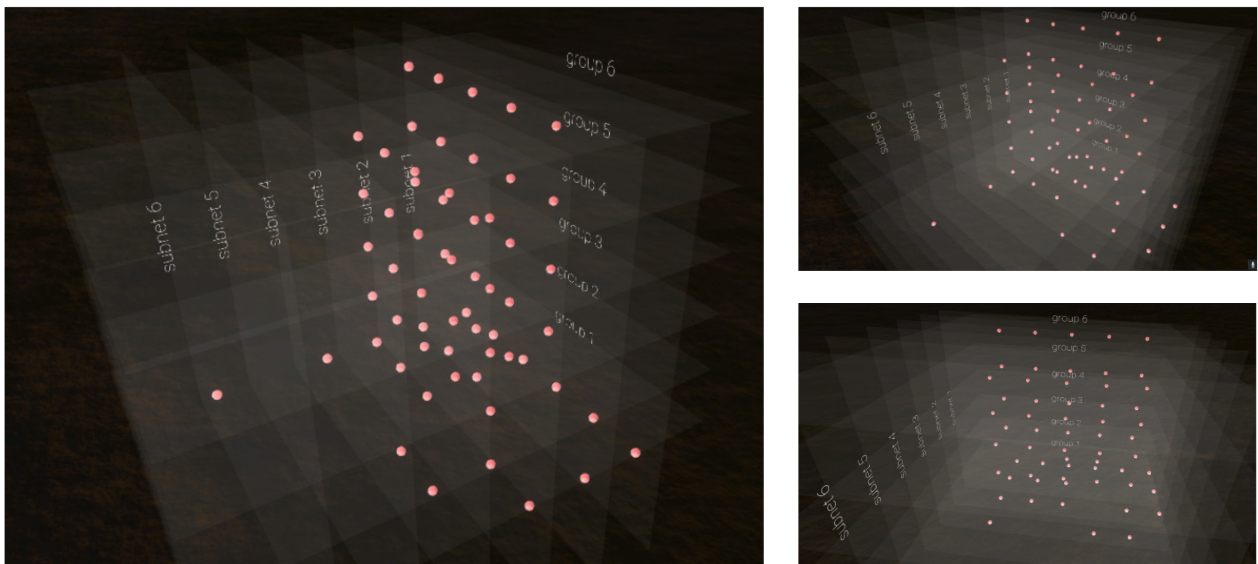


Figure 2-5: Display of a Blue Team’s Network Layout Where Entities’ Positions on XYZ Axes are Determined by: Z) The Group this Entity Belongs to (a Subnet); Y) Subgroup (a Functional Group in that Subnet: Servers, Networks Devices, Workstations); X) Entity’s Sequential (Arbitrary) Position in that Subgroup (for Example the Last Octet of its IP Address). (Source: Refs. [17], [18].)

2.4 EVALUATING DESIGN EFFECTIVENESS

The evaluation of design effectiveness (usefulness and usability) is the final phase of the human factors design process. Once again, users, playing the role of evaluation participants, are essential for meaningful testing. Designing an experiment to evaluate the effectiveness of a design with representative user tasks is the best way to confirm the fit between the user and their work with the visualization/interface/tool that was designed. Measures of human performance, situation awareness, and workload are the three most common metrics for evaluation.

2.4.1 Subjective versus Objective Human Performance Measures

Both subjective and objective measures have a role in evaluation. Subjective measures consist primarily of opinions gathered via questionnaires and interviews and are very helpful early on in the design process. They can point out missing features and bring attention to problems with visualization features, such as labelling, and mismatches between user expectations and feature implementation. However, subjective metrics are limiting and are prone to user biases. For instance, users may not positively rate designs with which they are unfamiliar. Objective measures of performance are typically gathered during an experiment in which a participant must perform a representative task or tasks and quantitative metric such as time and accuracy are recorded. Task analysis information from the analysis phase should provide specific user tasks with which to evaluate performance. Accuracy can be characterized in a number of ways to include a correctness score, number of corrects, number of errors, error rate, false alarm rate, deviations from optimum, percent correct, percent errors, ratio of number correct to number of errors, etc. Time can be characterized as reaction time to a stimulus, search time, marking speed, time to complete a task, etc.

2.4.2 Situation Awareness Measures

Situation Awareness (SA) is simply defined as “being aware of what is happening around you and understanding what that information means to you now and in the future.” [7]. As with human performance measures, there are both subjective SA measures (qualitative opinion data) as well as objective SA measures (quantitative performance data – typically, accuracy of responses to task-relevant probe questions). While self-report SA gathered with subjective SA measures can be useful, objective measures of SA are deemed as more helpful in determining the user’s actual awareness of the situation pertinent to performing a particular task [19]. Table 2-2 shows a partial list of both subjective and objective measures of SA.

Table 2-2: Measures of SA.

Subjective	Objective
Situation Awareness Rating Technique (SART)	Situation Awareness Global Assessment Technique (SAGAT)
Situation Awareness Subjective Workload Dominance (SA-SWORD)	Quantitative Analysis of SA (QUASA)
China Lake Situational Awareness	SA Analysis Tool (SAVANT)
Situation Awareness Behavioral Rating Scale (SABARS)	Situation Present Assessment Method (SPAM)
Mission Awareness Rating Scale (MARS)	Cranfield SA Scale (Cranfield-SAS)

2.4.3 Workload Measures

Workload indicates the level of work that a person is exerting while performing a task. Performance and workload are not negatively correlated (lower workload results in higher performance and higher workload results in lower performance) rather, the relationship between workload and performance is an inverted-U shape. Therefore, there is an optimum range of user workload that results in the highest performance. Too much workload leads to information overload and performance errors; not enough workload leads to boredom and performance errors. As with SA, there are a number of subjective workload measures and objective workload measures [7]. Subjective measures of workload include NASA's Task Load Index (TLX), Subjective Workload Assessment Technique (SWAT), and the Cooper-Harper Scale. Objective measures include physiological measures such as electro-encephalograms and performance measures including task errors and measures of space capacity using secondary tasks.

2.5 SUMMARY

The field of visual analytics focuses on analytical reasoning facilitated by interactive visual interfaces. These interfaces are essential to understanding the analytical techniques applied to data. Therefore, it is of utmost important to design those interfaces using human factors and the user-centered design process. Once the user or users of the visual analytics systems are defined, the phases of analysis, design, and evaluation as described in this chapter can help lead the design team to a system that is both useful and usable.

2.6 REFERENCES

- [1] International Ergonomics Association (IEA). What is Ergonomics? <https://iea.cc/what-is-ergonomics/> Accessed 10 Apr 2020.
- [2] California State University Long Beach. Department of Psychology. Option in Human Factors. <http://www.cla.csulb.edu/departments/psychology/ms-human-factors/> Accessed 10 Apr 2020.
- [3] Sanders, M., and McCormick, E. (1993). Human Factors in Engineering and Design, 7th Edition, McGraw Hill.
- [4] Salvendy, G. (Ed.) (2012). Handbook of Human Factors and Ergonomics, 4th Edition, Wiley.
- [5] Norman, D. (2013). The Design of Everyday Things, Basic Books.
- [6] Interaction Design Foundation. <https://www.interaction-design.org/> Accessed 10 Apr 2020.
- [7] Endsley, M., and Jones, D. (2012). Designing for Situation Awareness: An Approach to User-Centered Design, 2nd Edition, CRC Press, p. 13.
- [8] Roth, C. (13 Mar 2017). Why Henry Ford's Most Famous Quote is Dead Wrong. <https://www.entrepreneur.com/article/290410> Accessed 01 Apr 2020.
- [9] Eltin, L.S., Martin, C.G., Cantor, S.B., and Rubenstein, E.B. (5 June 1999). Influence of Data Display Formats on Physician Investigators' Decisions to Stop Clinical Trials: Prospective Trial with Repeated Measures. *BMJ.*; 318(7197): 1527-1531. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC27896/> Accessed 01 Apr 2020.
- [10] Nielsen, J. (24 Apr 1994, updated 15 Nov 2020). 10 Usability Heuristics for User Interface Design. <https://www.nngroup.com/articles/ten-usability-heuristics/>

- [11] Department of Defense Design Criteria Standard: Human Engineering, MIL-STD-1472F, (23 Aug 1999). http://everyspec.com/MIL-STD/MIL-STD-1400-1499/MIL-STD-1472F_208/ Accessed 01 Apr 2020.
- [12] Human Factors for Designers of Systems DEF STAN 00-25, Ministry of Defence Standard 00-250 (22 Oct 2012).
- [13] ISO. Ergonomics of Human-System Interaction – Part 210: Human-Centred Design for Interactive Systems ISO 9241-210:2019 (Jul 2019). <https://www.iso.org/standard/77520.html> Accessed 10 Apr 2020.
- [14] Furman, S., Theofanos, M., and Wald, H. (Apr 2014). Human Engineering Design Criteria Standards Part 1: Project Introduction and Existing Standards. DHS S&T TSD Standards Project <https://nvlpubs.nist.gov/nistpubs/ir/2014/NIST.IR.7889.pdf> Accessed 01 Apr 2020.
- [15] Kullman, K., Ben-Asher, N., and Sample, C. (2019). Operator Impressions of 3D Visualizations for Cybersecurity Analysts. ECCWS 2019 18th European Conference on Cyber Warfare and Security, Coimbra, 2019.
- [16] Kullman, K., Cowley, J., and Ben-Asher, N. (2018). Enhancing Cyber Defense Situational Awareness Using 3D Visualizations. 13th International Conference on Cyber Warfare and Security, Washington, DC, 2018.
- [17] Kullman, K., Ryan, M., and Trossbach, L. (2019). VR/MR Supporting the Future of Defensive Cyber Operations. 14th IFAC Symposium on Analysis Design and Evaluation of Human Machine Systems, Tallinn, 2019.
- [18] Kullman, K., Buchanan, L., Komlodi, A., and Engel, D. (2020). Mental Model Mapping Method for Cybersecurity. 22nd International Conference on Human-Computer Interaction, Copenhagen, 2020.
- [19] Gawron, V. (2019). Human Performance and Situation Awareness Measures, 3rd Edition, CRC Press.



Chapter 3 – INFORMATION VISUALIZATION AND VISUAL ANALYTICS FOR THE MARITIME DOMAIN¹

Elena Camossi
NATO STO CMRE
ITALY

Cyril Ray
Ecole Navale
FRANCE

Ethem Arkin
ASELSAN
TURKEY

3.1 INTRODUCTION

The North Atlantic Treaty Organization (NATO) Research Task Group on exploratory visual analytics works to promote the research and deployment of visual analytics techniques among NATO member and partner nations across a broad range of NATO application areas.

In the maritime domain, visualization techniques are key components for a better situational awareness. A (good) adapted visualization may provide many benefits by helping expert analysts understand the maritime situation (especially anomalies and cyber threats) while reducing their cognitive load. This later point is particularly important regarding the data deluge experts have to cope with nowadays. The data analysis upstream is also a key factor in developing good visualization techniques. It helps filtering, organizing, and presenting not only data, but structured information with a higher level of semantics.

Beyond the development and application of different visual analytics tools to support maritime domain analysis done in the task group, this chapter aims to report on typical data and data visualizations presented in the literature. The maritime domain is made of a large variety of data. However, maritime situational awareness often relies on visualizing navigation data, mainly collected in real time, but also historical data collected, stored, and analyzed offline, in order to reach information superiority.

In order to investigate innovative Interactive Visualizations (IVIS) and Visual Analytics (VA) for maritime data, the IST-141/RTG 66 has accomplished several activities, detailed in the rest of the chapter. To identify potential innovations and novel research avenues, maritime datasets have been shared among the group members to encourage the cross fertilization of ideas. Another expected objective was the transfer of technologies among different disciplines. Section 3.2 describes these datasets. A state of the art of the recent results of IVIS and VA in the maritime domain is presented in Section 3.3. Maritime visualization applications for operational use and research oriented Visual Analytics tools, described in Section 3.4, have been demonstrated, and realistic maritime use cases have been discussed with the group to identify capability gaps and research problems driving the development of innovative techniques and research advancements for IVIS and VA. Section 3.5 concludes the chapter.

3.2 MARITIME DATASETS

CMRE (NATO) and NARI ENSAM have shared with the RTG members a series of maritime datasets collected from their facilities, in order to encourage cross fertilization among domains and the development of innovative maritime IVIS and VA.

¹ An extended version of the content presented in this chapter appeared as a CMRE report and is available upon request. Reference number is CMRE-MR-2019-28 (NATO UNCLASSIFIED Releasable to Sweden).

In the recent years, the open literature addressing maritime applications has extensively exploited the data generated by the *Automatic Identification System* (AIS)², a collaborative collision avoidance system that broadcasts messages via radio communication. AIS messages give information on ships kinematic (e.g., current position, speed, expected destination and time of arrival) and provide additional ships meta-information (e.g., name, type, international unique identifiers). The European legislation obliges several types of ships to broadcast AIS messages, including: ships of 300 gross tonnage and upwards in international voyages; cargoes of 500 gross tonnage and upwards navigating outside international waters and passenger vessels; and, more recently, smaller fishing vessels. Raw AIS messages comply with International Telecommunication Union Recommendation, Series M (ITU-R.M) 1371-5 and NMEA (National Marine Electronic Association) 4.0 standards and are differentiated in 27 types of messages. AIS messages can be collected by a network made by coastal and satellite receivers. *Terrestrial AIS* (T-AIS) messages from coastal receivers are characterized by high frequency but limited coverage, while *Satellite AIS* (S-AIS) messages collected by satellite receivers that can pick up messages in the open sea, far away from the coastline, are sparser but have a larger coverage than T-AIS.

CMRE shared with the group four datasets of decoded AIS messages collected during the execution of the Decision Knowledge Operational Effectiveness (DKOE) program, through the CMRE's terrestrial receiver (*Castellana*, located in the port of La Spezia, Italy), and through *AISHub*³, an open network of AIS receivers. These datasets contain the following data, and are distributed according to the security classification as specified in brackets:

- One week of T-AIS messages by the receiver *Castellana*, transmitted by vessels travelling in the gulf of La Spezia and on the North of the Tyrrhenian Sea during the 1st week of August 2016 (NATO UNCLASSIFIED).
- 323 MB of T-AIS messages, collected by the receiver *Castellana*, transmitted by vessels travelling in the gulf of La Spezia and on the North of the Tyrrhenian Sea along August 2015 (NATO UNCLASSIFIED, Releasable to Sweden).
- 74 MB of T-AIS collected through the *AISHub* AIS network, transmitted by vessels travelling in the Aegean Sea during the period 1 – 15 August 2015 (PUBLIC RELEASE).
- 712 MB of T-AIS collected through the *AISHub* AIS network, transmitted by vessels travelling in the Baltic Sea during August 2016 (PUBLIC RELEASE).

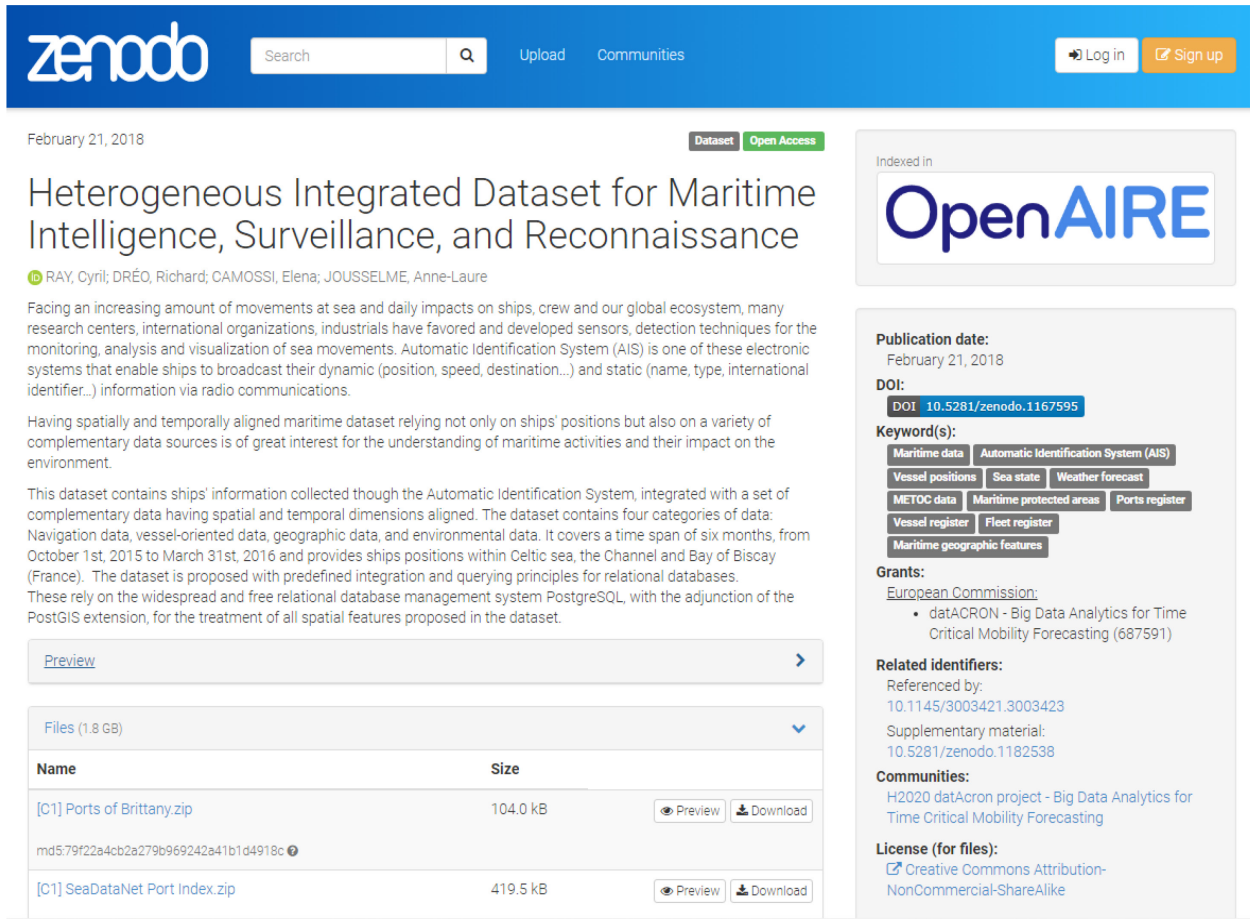
In addition to these datasets, NARI ENSAM has shared with the RTG members an open maritime dataset that has been created for the evaluation phase of the H2020 project “Big Data Analytics for Time Critical Mobility Forecasting” (datAcron) [1], [2]. This dataset, which is available for download through the open science platform *Zenodo*⁴ (see Figure 3-1) addresses a fishing use case and contains AIS messages collected by the NARI's receiver, which is positioned in the area of Brest, Brittany. These data are integrated with a set of complementary data, aligned in time and space. The dataset contains four categories of data:

- **Navigation data:** raw AIS messages, AIS status, country codes, information on receptors like location and theoretical coverage;
- **Vessels' registers;**
- **Geographic data:** coastline, ports, seas, fishing areas; and
- **Environmental data:** protected areas, weather data, ocean models, sea state, and fishing areas.

² www.navcen.uscg.gov/?pageName=AISmain

³ www.aishub.net

⁴ <https://doi.org/10.5281/zenodo.1167595>



The screenshot shows the Zenodo interface for a dataset. At the top, there is a search bar and navigation links for 'Upload' and 'Communities'. The dataset title is 'Heterogeneous Integrated Dataset for Maritime Intelligence, Surveillance, and Reconnaissance', published on February 21, 2018. The authors listed are RAY, Cyril; DRÉO, Richard; CAMOSSO, Elena; and JOUSSELME, Anne-Laure. The dataset is described as containing ships' information from the Automatic Identification System (AIS), integrated with other data like navigation and environmental data. It covers the period from October 1st, 2015, to March 31st, 2016, in the Celtic Sea, Channel, and Bay of Biscay. The dataset is available in two files: '[C1] Ports of Brittany.zip' (104.0 kB) and '[C1] SeaDataNet Port Index.zip' (419.5 kB). The page also features a list of keywords such as 'Maritime data', 'Automatic Identification System (AIS)', 'Vessel positions', 'Sea state', 'Weather forecast', 'METOC data', 'Maritime protected areas', 'Ports register', 'Vessel register', 'Fleet register', and 'Maritime geographic features'. It mentions grants from the European Commission, specifically the datACRON project. The license is Creative Commons Attribution-NonCommercial-ShareAlike.

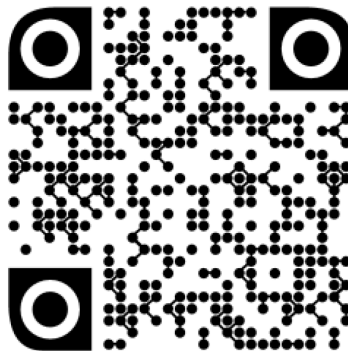


Figure 3-1: Maritime Open Data Available for Research from Zenodo <https://doi.org/10.5281/zenodo.1167595> [1], [2].

The dataset covers a time span of six months, from October 1st, 2015 to March 31st, 2016 and provides ships positions in the Celtic Sea, the Channel and Bay of Biscay (France). The dataset is proposed with predefined integration commands and queries for relational databases. These rely on the widespread and free relational database management system PostgreSQL, with the adjunction of the PostGIS extension for the treatment of all spatial features proposed in the dataset.

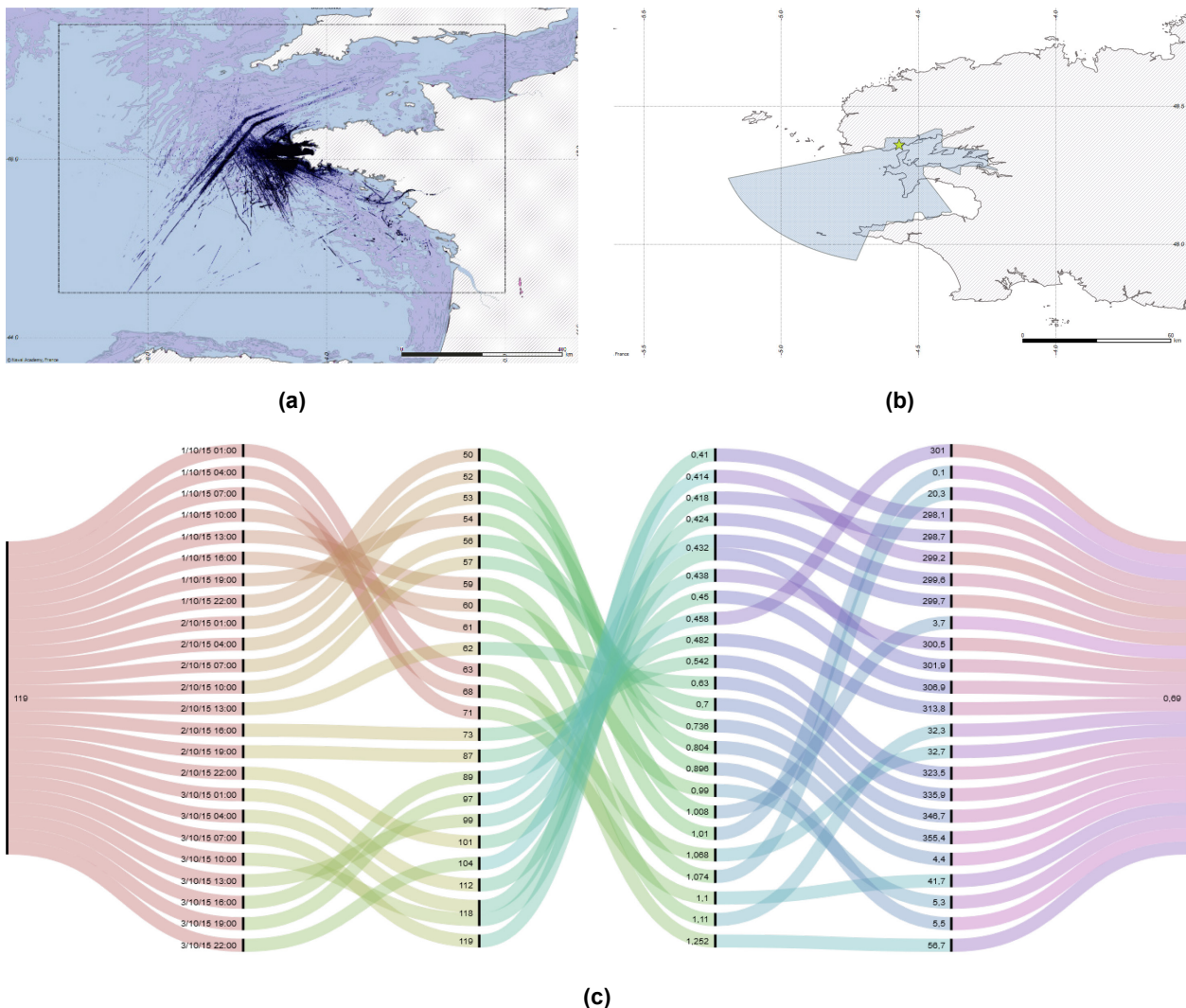


Figure 3-2: Example of Data Included in the Dataset at Refs. [1], [2] [CC-BY 4.0]. (a) NARI AIS data with (b) the theoretical coverage of the NARI AIS receiver. (c) Alluvial diagram illustrating the correlation of ocean condition variables along 3 days (from left to right: depth, date, mean wave length, height of wind and swell waves, wave mean direction, sea surface height).

3.3 STATE OF THE ART ON MARITIME VISUALIZATION

The recent literature has extensively investigated the use of visualization and novel visual analytics approaches for maritime applications.

Earlier work investigated the application of interactive visualizations and exploratory visual analytics techniques for *maritime security*. These works mainly addressed traffic monitoring and the understanding of traffic dynamics.

The following work investigated the development of visual analytics for *pattern extraction and detection*. These works, usually integrating also interactive visualization capabilities, have been driven by the interest for developing automated early warning systems for *maritime situational awareness* and *maritime anomaly detection*. The developed visual analytics are usually of three types: data driven, where a model of normal behavior is built from historical data and compared against real-time data to detect outliers; *rule driven*,

where events of interest are defined as patterns, sometime interactively, and real-time data are evaluated against such patterns to detect events and trigger the corresponding warnings; and *hybrid*, which combine the two approaches.

3.3.1 Interactive Visualization of Vessel Traffic for Monitoring and Exploratory Analysis

The approaches presented in the literature investigate the visualization of maritime data in order to interactively explore their spatial distribution and their temporal evolution, or to facilitate their comparison by combining aggregation and filtering capabilities. The use of map-based visualizations prevails, occasionally complemented by statistical graphs of data aggregates. Recently, graph-based visualization of maritime data has also emerged. Graph-based visualizations allow combining aggregation statistics with visualization of traffic dynamics.

Vatin and Napoli [3] investigate the use of geographical visualizations for maritime traffic control. The study concludes that an effective visualization for maritime traffic control should take into consideration both the space and time dimensions, and be interactive, leaving the user free to choose which are the relevant attributes to visualize. This work focuses on maritime risk factors such as the ship's behavior, the geographical area (i.e., dangerous or not), and the contextual situation (e.g., ship type, visibility).

Malik et al. [4] propose a visualization for maritime risk assessment that may be used for optimal resource allocation for Search and Rescue (SAR). The allocation is based on the risk assessment of the monitored areas, which relies on historical data of past SAR operations. The visualization dashboard combines a map showing the spatial distribution of SAR operations with their temporal distribution and the type of associated distress. In order to support the user in the analysis, temporal filtering and different temporal aggregation levels are also provided.

In Ref. [5], Cazzanti et al. develop an operative tool (the *Maritime Patterns-of-Life Information Service*, MPoLIS) that automatically generates and visualizes traffic port statistics from AIS data streams, which are continuously processed according to a big data software architecture. The statistics are published in an interactive dashboard developed with Tableau. Different visualizations are available, including line charts that show the temporal evolution of the data, and histograms, which enable the comparison of alike data aggregates (e.g., unique port visits in ports of the same countries). As discussed later in the chapter (cf. Section 3.4.1), MPoLIS successfully demonstrates the use of an off-the-shelf visualization software for the rapid prototyping of automated statistics in maritime.

Andrienko and Andrienko [6] use, among other visualizations, time bars and temporal histograms for the analysis of the speed of vessel trajectories (see Figure 3-3).

Last et al. [7] propose a visualization tool for improved MSA that fuses AIS data with radar data on a radar-like visualization. The proposed visual encoding combines different layers of information available from AIS (i.e., type of moving object, movement, and dimension of the vessel), and is consistent with encodings and standards defined by the International Maritime Organization, IMO (see Figure 3-4).

In Ref. [8], Last presents an interactive visualization for vessel movement prediction where position estimations are calculated on the basis of historical AIS data. Polygonal lines representing the background model trajectories and predicted vessel trajectories are overlaid on a map allowing the visual inspection of the interesting targets. The resulting visualization is shown in Figure 3-5. Interestingly, the visual clutter produced by overlapping vessel trajectories in high-density areas is not smoothed, like in density maps.

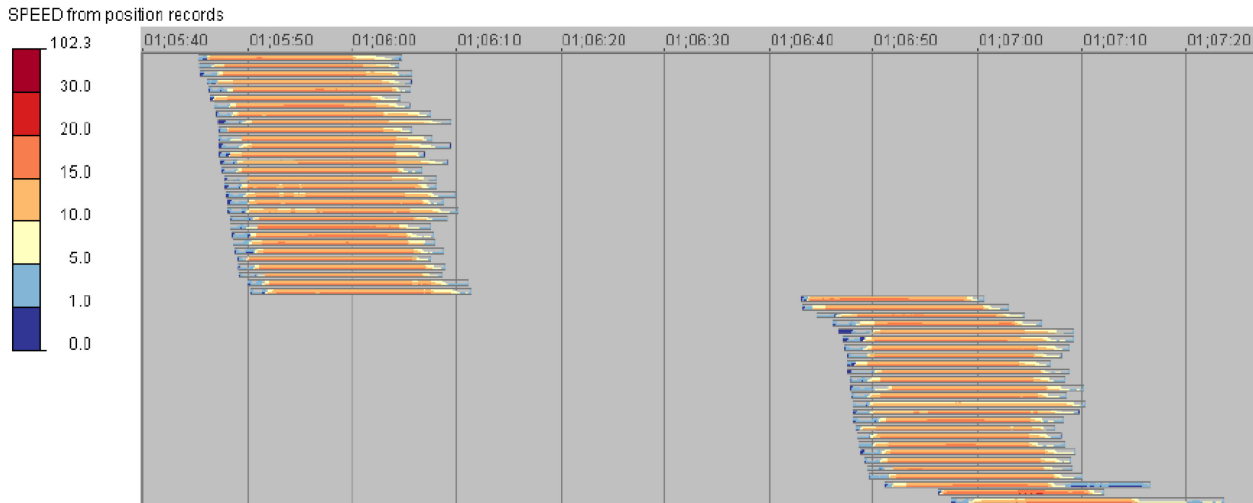


Figure 3-3: Time Bars for the Analysis of Temporal Variation of Speed in Vessel Trajectories [6] (©Andrienko and Andrienko 2013).

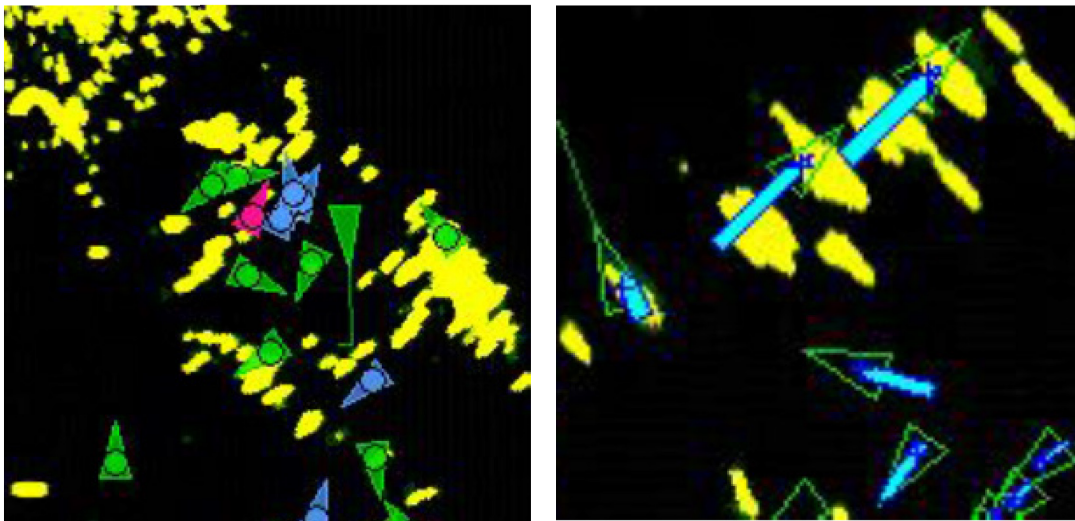


Figure 3-4: Visualization Integrating AIS Information (Triangular Icons) and Radar Contacts (Image in Yellow). AIS encodings include the type of the vessel, its movement and size. On the left: vessel types (color) and movement (dotted, non-moving vessels, vs plain icons). On the right: the size of the vessel (blue bar) [7] (© Vaclav Skala – UNION Agency).

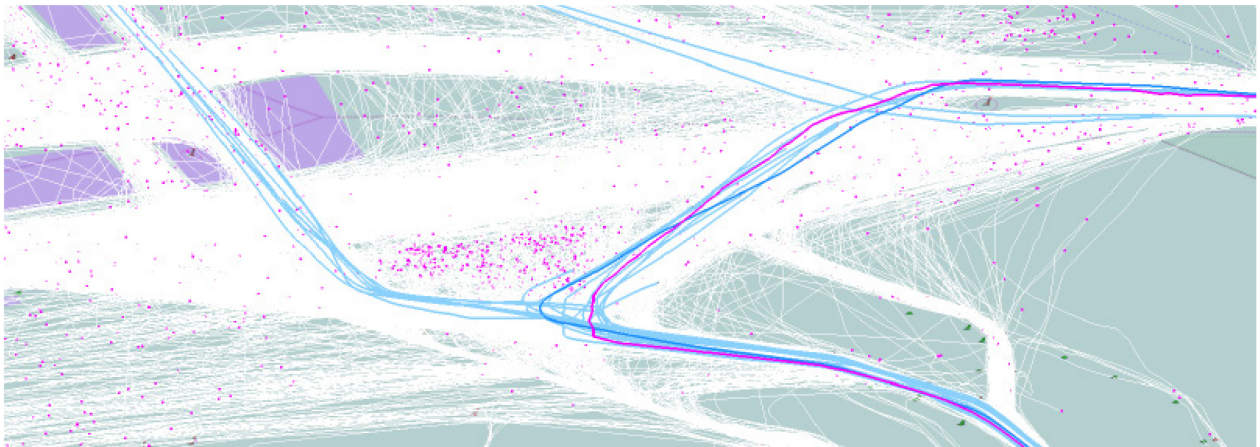


Figure 3-5: Interactive Visualization of Vessel Trajectory for Prediction [8]. The predicted trajectory (in magenta) is overlaid to the ground truth (in dark blue) and to the historical data (in light blue) (© Philipp Last 2016).

The overall ship movements can be modelled with a graph, whose nodes depict the stop locations (e.g., ports, mooring areas), and edges link stops [9]. Therefore, ship trajectories can be grouped along nodes and edges, and aggregates can be statistically analyzed. This node-edge model provides a useful data structure to analyze and visualize differently a very large set of navigation data. By using such an approach data can be manipulated, queried, analyzed, and visualized, creating analytics on maritime transportation networks, or to visualize ships' life cycle using a graph-based visualization, as illustrated in Figure 3-6.

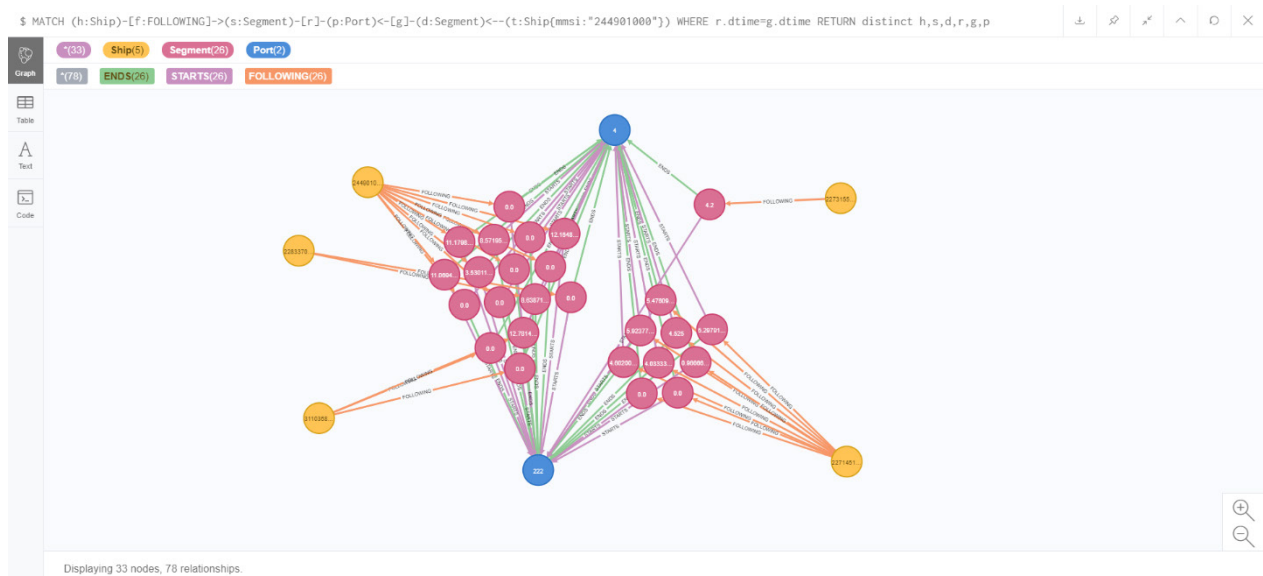


Figure 3-6: Detection of Vessel Meeting Points: For a Given Ship, the Graph Illustrates the List of Visited Ports and the List of Other Ships that Were in the Same Port at the Same Time (© Itani and Ray, 2019).

3.3.2 Visual Analytics for Maritime Pattern Detection

Visual analytics, which combine effective maritime data visualizations with data analytics, has a variety of applications. One of the uses discussed in the literature is movement patterns search and matching, which,

combined with data filtering, enables to reduce data deluge and help the user focus on relevant data characteristics. This approach defines movement patterns empirically, for instance using function based definition or rules. Movement patterns may also be extracted from data, for instance using data clustering, aggregation and filtering techniques. Geographical features are exploited to support the definition and the visualization of patterns. Different aggregation techniques can be combined, sometime with the use of semantics, to support data exploration. Other works specifically address anomalous or inconsistent pattern detection. For instance, the detection or forecasting of close encounters is applied to maritime safety for vessel collision identification and prevention.

Enguehard et al. [10] present a geographical visualization tool for the interactive analysis of vessel trajectories reconstructed from Vessel Monitoring System (VMS) data, another positional self-reporting system. The tool supports the definition of movement pattern signatures based on fractal and velocity properties, which adapts well to the definition of fishing patterns and to vessel activity classification. Movement signatures are combined with temporal filtering. The resulting specification produces a hybrid spatio-temporal pattern, which is used to filter the data. By reducing the amount of data shown to the analyst, filtering helps the user focus on interesting events.

Andrienko and Andrienko [11] illustrate visual analytics techniques applied to the analysis of maritime data (specifically, AIS messages). They demonstrate the use of clustering and discrete aggregation to the discovery of traffic lanes and flows (see Figure 3-7).

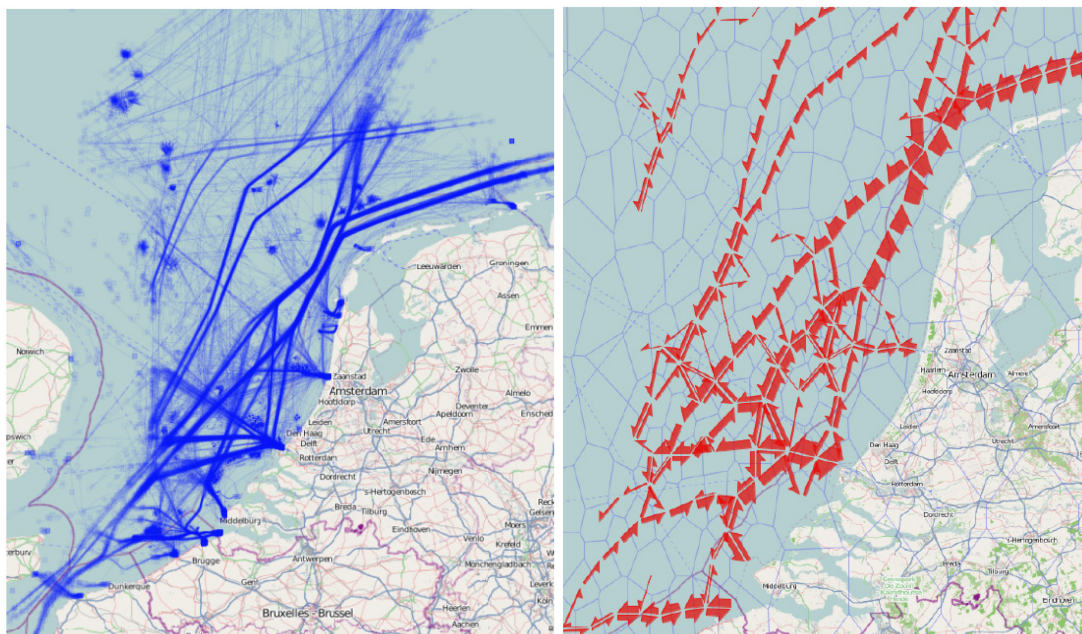


Figure 3-7: Clustering and Discrete Aggregation to Identify Vessel Traffic Lanes and Flows [11] (© Andrienko and Andrienko, 2011).

The Dutch project POSEIDON in 2007 [12] developed an innovative online visualization and analysis tool to support interactive visual hypothesis testing for monitoring of vessel traffic in coastal areas and anomaly detection. Vessel trajectories are reconstructed from live data generated by heterogeneous sensors, then semantically enriched with geo-semantics and additional information gathered from the web. The resulting maritime traffic and event ontology may be annotated for further reasoning and queried by experts through a visual analytics module that combines the search of spatio-temporal patterns of interest with analysis capabilities. The visual analytics module in Ref. [12] integrates in a dashboard attribute filtering, interactive Trajectory Contingency Tables (TCT) and a spatial overview of the vessel trajectories, supporting the

exploration of the vessel traffic knowledge base. Interactive TCT are used to highlight existing correlations between vessel attributes. For example, the user can highlight the changes over *time* for different *vessel types*, and selected vessels trajectories are depicted in a geographical map, all in the same dashboard.

More recently (see for instance Refs. [13], [14]), the visual analytics module of the POSEIDON Maritime Security System has been enhanced with interactive density fields, used to detect vessel trajectory patterns, specifically to identify stopping areas like anchor zones and sea lanes, to create custom statistics (e.g., accident risk maps by vessel type), and to identify temporal trends. The aggregation of two density fields representing the historical and the live traffic is used to facilitate the detection of inconsistencies.

Andrienko et al. [15] use continuous density fields and discrete aggregations of vessel trajectory flows in combination with geospatial analysis functions and time masking filtering to analyze vessel traffic, specifically to analyze *near-location* events, which potentially increase the risk of collision, in the area of Brest. Density fields, combined with time mask filtering, enable the user to focus the analysis on near collision events identified empirically using the geospatial analysis capability of the tool (Figure 3-8 top). The same data are also visually analyzed using a discrete aggregation that shows the direction of the traffic flows and their volumes (Figure 3-8, bottom).

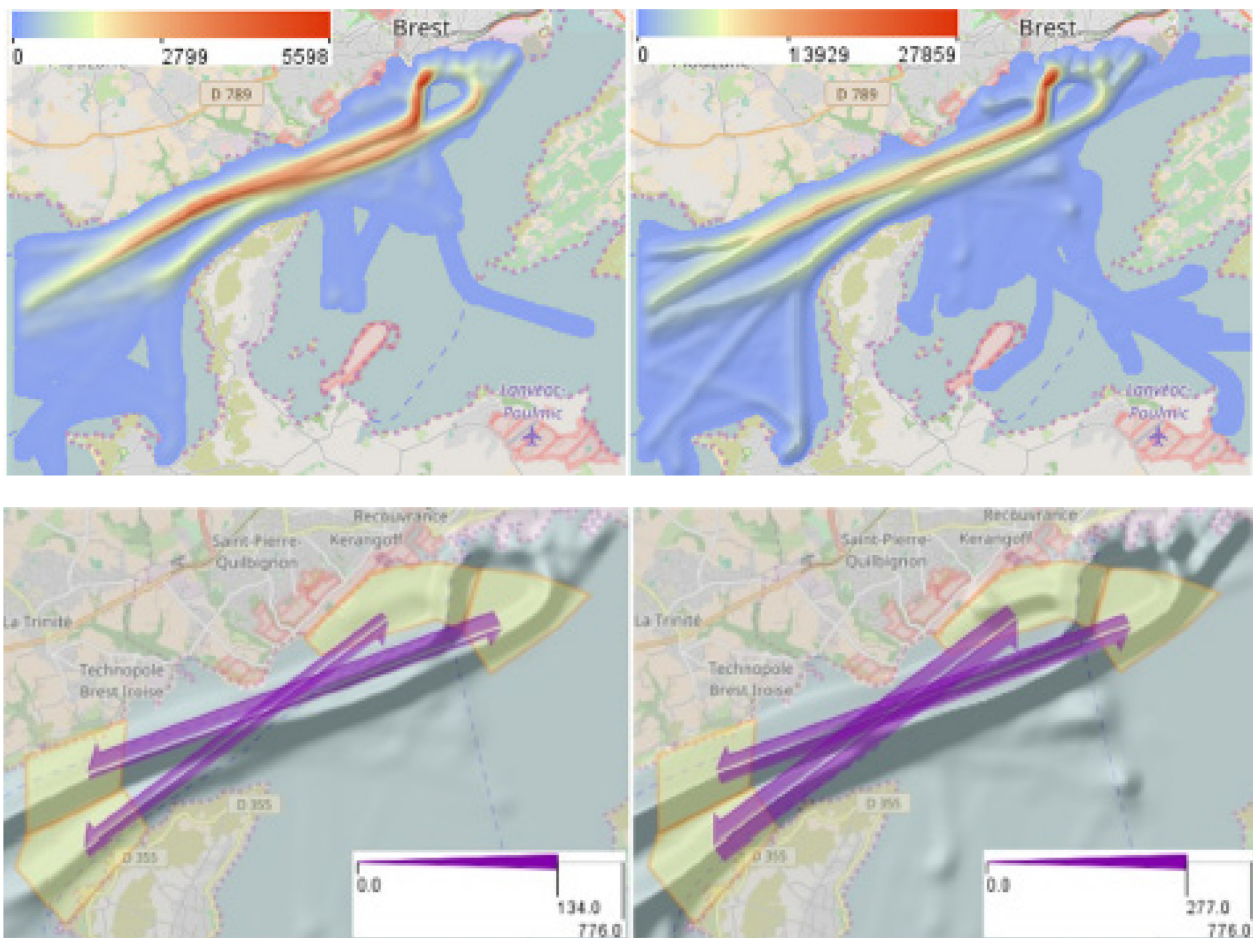


Figure 3-8: Analysis of Near Collision Events in the Port of Brest. Top Right: density fields of vessel traffic when near collision events at rush hours. Top Left: Density fields of vessel traffic when no near collision events occur. Bottom left: discrete aggregation showing vessel traffic flows when near collision events are detected. Bottom right: Vessel traffic flow during non-pick hours [15] (CC-BY-NC-ND 4.0).

Scheepens et al. [16] propose a *contour* based visualization of vessel movement predictions. The prediction model is based on historical vessel trajectories and weather data. The predicted position of a vessel is visualized by the contours of the temporal probability density fields, and fading is used to show increasing future instants, reducing the user's information overload. Density fields and contours are used in combination to visualize the probability of interactions between vessels, to identify potential collisions and encounters for maritime security applications (e.g., smuggling, piracy). The higher the density of the fields, the higher is the probability of interaction between vessels.

Defence Research and Development Canada (DRDC) [17], [18], [19] developed a service-based maritime visualization for the exploratory analysis of vessel traffic that is interactive and customizable. The service includes automated visual analytics capabilities for the detection of events and anomalies. Normal maritime behaviors (e.g., commercial routes, fishing areas) can be visualized and exploited for surveillance and anomaly detection (e.g., close encounters and *rendez-vous*, package drop-off). The interface combines multiple widgets that offer different interactive visualizations. The geographical view supports animated maps, enabling the user to interactively visualize the vessel movements in an area of interest to highlight anomalies like close encounters. Magnets grids enable to interactively visualize outliers based on vessel properties. Multi-timelines allow visually comparing temporal events.

Riveiro et al. [20] have developed an interactive visual analytics tool for the detection of maritime anomalies that integrates the geographical visualization of a normalcy model built from historical AIS data, which uses Self Organizing Maps (SOM), and a Gaussian Mixture Model. SOM give a visual representation of the clustered data, facilitating the tuning of the model parameters. In Ref. [21], the same authors discuss the effectiveness of interactively visualizing normal vessel behavior for maritime anomaly detection. The tool in Ref. [20] was also extended to support the detection of early warnings, or anomalies, detected according to rules specified by the user (i.e., incongruences in the reported messages; kinematic anomalies on positions, speed, heading; kinematic anomalies involving multiple objects for piracy, smuggling, anomalies combining speed, position and headings). Functionalities for querying the historical data are also supported [22].

3.3.3 Three-Dimensional Visualization of Maritime Pattern and Mobility

Few works exploit the use of three-dimensional visualizations in the maritime domain, because of the increased complexity of the resulting visualization, which can affect negatively the user experience. Domain-agnostic three-dimensional visualizations are used to support the comparison of single vessel dynamics parameters. The *space-time cube* visualization is effective for the analysis of maritime events, because it enables the contemporary visualization of spatial and temporal characteristics. Immersive virtual reality models, which are used in training, may be more effective than two dimensional maps to support the user (and vessel) orienteering.

Andrienko and Andrienko [6], beside time bars and temporal histograms, use spatio-temporal visualizations for the analysis of the speed of vessel trajectories, as shown in Figure 3-9.

The same authors illustrate in [11] the use of a space-time cube, applied to the detection of shift proximity and drifting events in vessel trajectories (see Figure 3-10).

Etienne et al. [23] present an interactive geo-visual analysis tool to detect spatio-temporal outliers in maritime trajectories. A model of normal behavior is constructed, aggregating spatio-temporal statistics of historical maritime trajectories, and used for comparison. As illustrated in Figure 3-11, outliers are visualized in a space-time cube to support both temporal and spatial anomalies. When the numbers of trajectories increases, the authors claim that analysis techniques based on clustering and data mining could improve the cognitive potential of the space-time cube helping the user focus on the most salient patterns and facilitating the visual identification of anomalies.

The time and mental effort required for understanding of two dimensional maps has severe consequences in areas where time for analysis of the situation is crucial. The strength of three-dimensional maps lies in particular in the perspective representation that mimics the way human perceive the world. For instance, Ref. [24] uses three-dimensional immersive visualization showing the point of view of navigating ships.

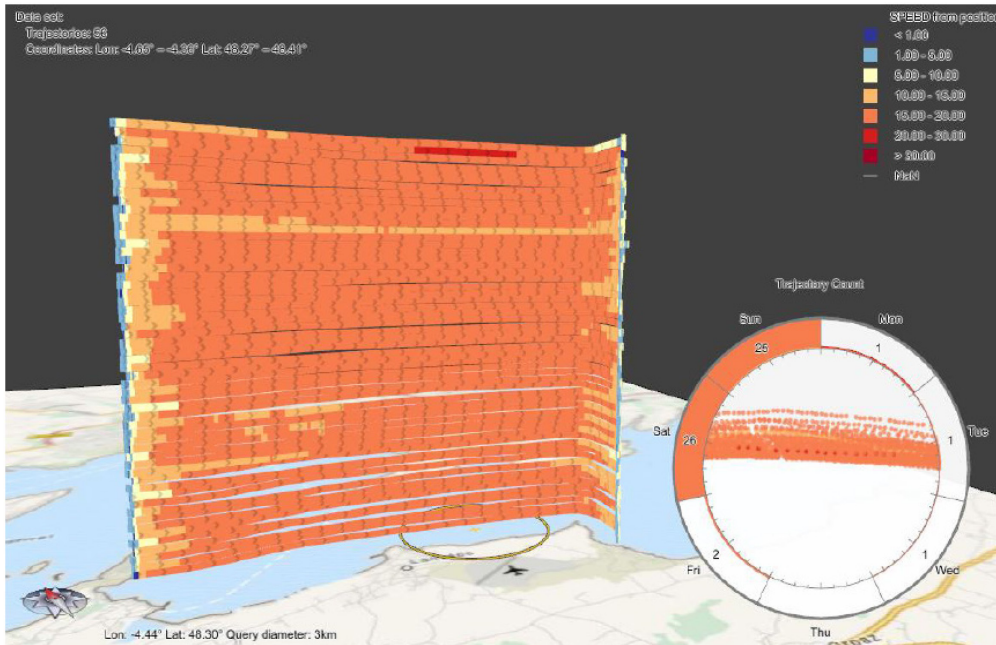


Figure 3-9: Spatio-Temporal Visualization for the Analysis of Speed Variations in Vessel Trajectories [6] (© Andrienko and Andrienko, 2013).

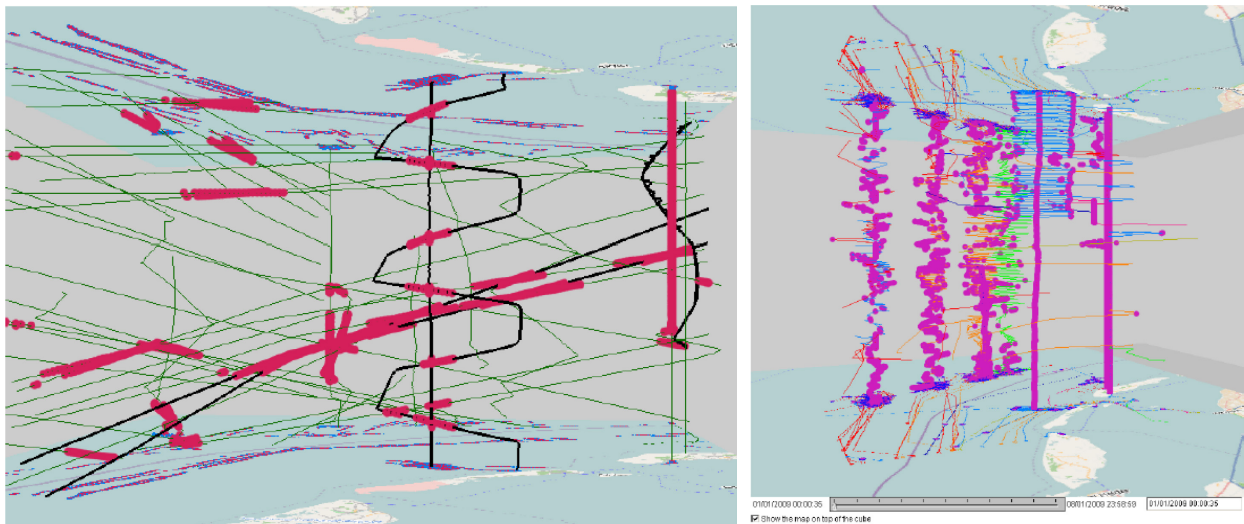


Figure 3-10: Space-Time Cube for Vessel Event Detection: Shift Proximity (Left) and Drifting (Right) [11] (© Andrienko and Andrienko, 2011).

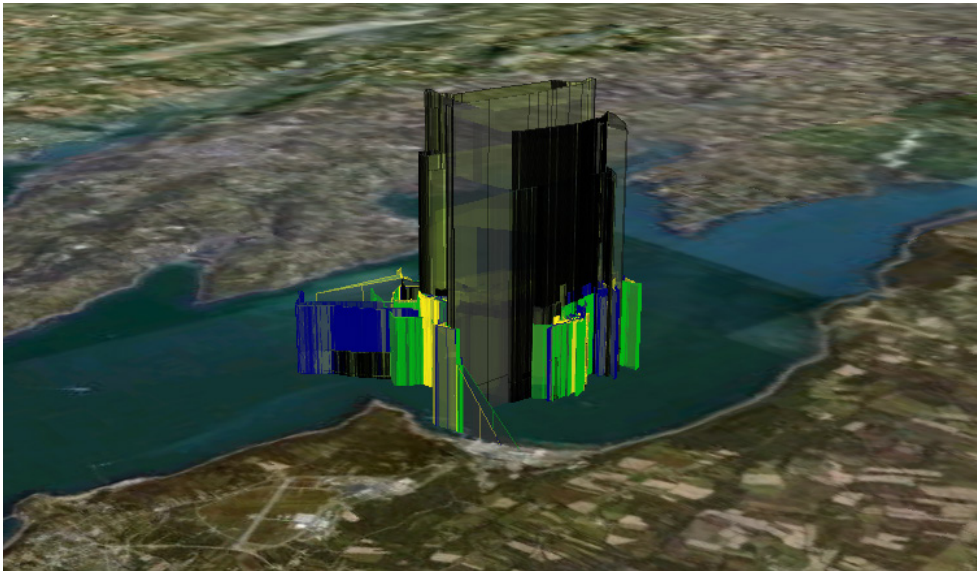


Figure 3-11: Space-Time Cube for the Visualization of Outliers in Vessel Traffic [23]
(© Etienne, Ray and McArdle, 2011).

3.4 EXPERIMENTS ON MARITIME INTERACTIVE VISUALIZATIONS AND VISUAL ANALYTICS

3.4.1 Maritime Patterns-of-Life Information Service (MPoLIS)

CMRE developed the Maritime Patterns-of-Life Information Service (MPoLIS) [5]. MPoLIS is a web service that automatically processes AIS data and generates different interactive visualizations of *vessel visits* in *port areas*. MPoLIS aggregates vessel positions according to country and time ranges (months, years), and produces statistics like total counts of unique vessel visits (by port and by vessel type); counts of vessel visits (by port and by vessel flag); time series of monthly vessel counts.

The interface of MPoLIS includes an interactive dashboard, which has been implemented using Tableau Online. Tableau⁵ is software platform for developing interactive visualizations and exploratory visual analytics. MPoLIS leverages Tableau Online to compute statistics on the fly, based on intermediate analysis results. This feature was extremely helpful during the development of MPoLIS, because it enabled to experiment and evaluate different visualization and data aggregations, facilitating the application of an agile approach for capturing the user requirements as well as favoring the user uptake.

Figure 3-12 (left) shows the MPoLIS Dashboard. Figure 3-12 shows statistics on Italian ports. On the left side of the dashboard, the monthly vessel visits are visualized through line charts, broken out by port, ship type, and state flag. On the right side, bar charts enable the comparison of the total vessel counts over a selectable range of dates. The user can select one or more ports, and the statistics are automatically drilled-down. Similarly, other statistics can be generated to investigate the vessel traffic per country, or region (e.g., Mediterranean). The user may save the screenshots of the dashboard to image files, download the underlying data and save them in ASCII format. The user can also move from the Tableau-based dashboard to the density map of the vessel traffic generated from the underlying AIS data (cf. Figure 3-12 (right)). Different map layers carry information on different vessel types. The maps are hosted on a separate map server (GeoServer) on the cloud. Both maps and statistics are updated on a monthly basis.

⁵ www.tableau.com

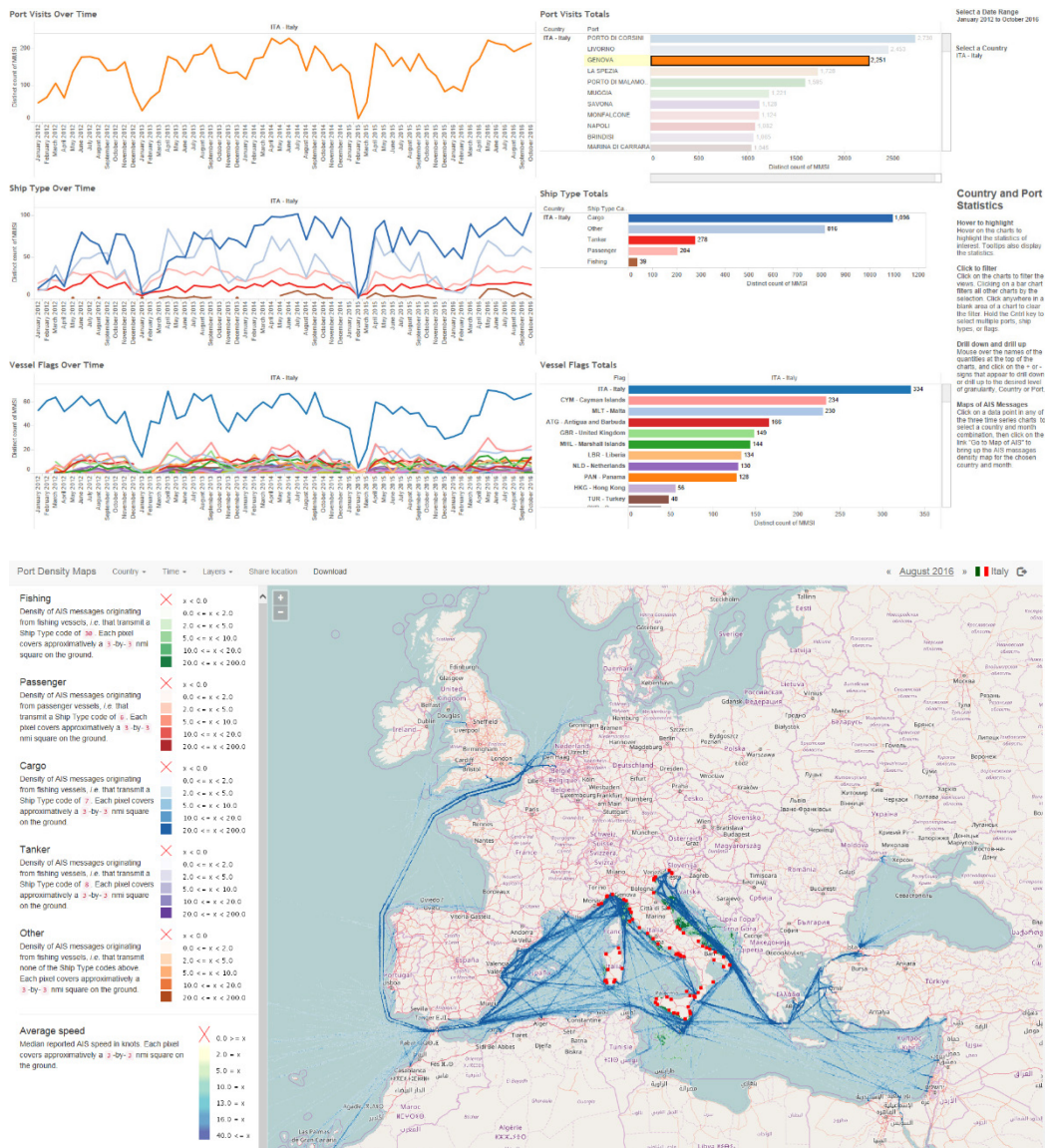


Figure 3-12: MPoLIS Interface, Showing Vessel Traffic for Italian Ports [5]. Top: MPoLIS Dashboard, showing port visits over time (top); port visits by ship type (middle); port visits by vessel flag (bottom). Bottom: density map of unique visits to Italian ports (Courtesy ©IEEE, 2016).

3.4.2 A Visual Analytics Experience for Naval Command and Control Application: Case Study on the Turkish Straits

The VATOZ[®] Naval Mission Management Software is a Command and Control (C2) application that has been developed by ASELSAN to support maritime domain awareness of Turkish Navy Base Defence Operations Centers and Turkish Patrol Boats. The application provides situational awareness for specific areas and shorelines to Navy operators. ASELSAN has implemented a case study to improve the situational awareness on the Turkish straits, which are an important maritime domain for Turkey. C2 applications are the essential parts of the defence systems used in these critical fields, because most decision-making processes and activities are performed by using such applications. Exploratory visual analytics methods offered an excellent opportunity to improve situational awareness. As such, a visual analytics extension for VATOZ[®] has been developed.

Formerly collected data from the Turkish straits domain have been stored in a data warehouse. Data analysis is performed to trace information from the AIS systems into the C2 system. Traces information is timestamped, then data are filtered and clustered. In order to support the efficiency of these functions and of metric calculation, the tool *Elasticsearch*⁶ has been evaluated because it offers indexing capabilities. *Elasticsearch* is a full-text search engine and data-analysis tool that supports efficient data storage and search. The *Logstash* tool included in the Elastic toolkit records transmitted data, which can be sorted appropriately and directed to *Elasticsearch*. The *Logstash* supports data received over standard input, file, or network. *Elasticsearch* can perform data-analysis operations on the stored data. Search-based analytics leverage indexing to improve search efficiency. Additionally, *Elasticsearch* aggregates data, clustering them according to given criteria. Clustered data are presented in metrics or buckets. The metrics provide statistical values such as total, mean, and maximum of the clusters, while the buckets provide the grouped information within the clusters.

Figure 3-13 and Figure 3-14 show typical visualizations for non-geographical data. In Figure 3-13, pie charts show statistics on vessel types and flags, in a given amount of time.

In Figure 3-14, line charts show the variation of vessel types and flags, along the same period of time.

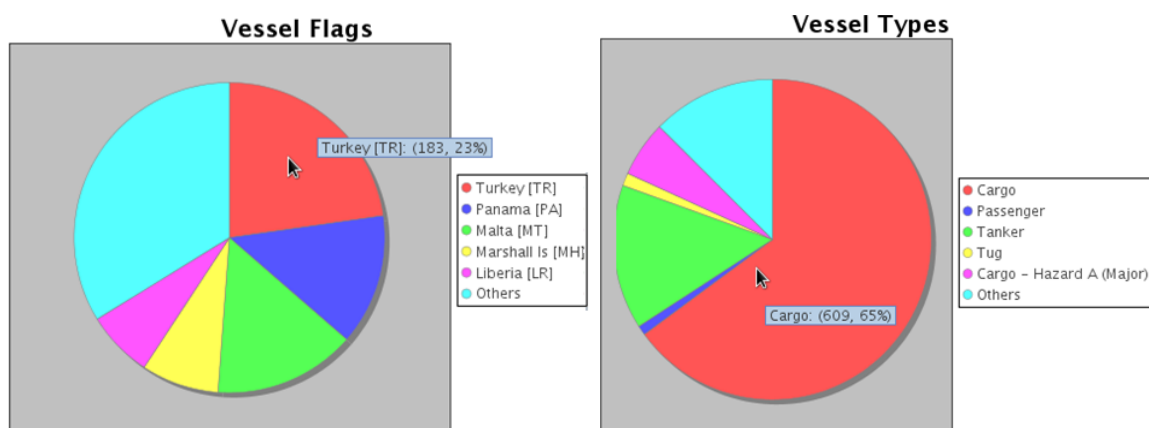


Figure 3-13: VATOZ® Visualizations. Pie charts of vessel type and flag density (©ASELSAN 2019).

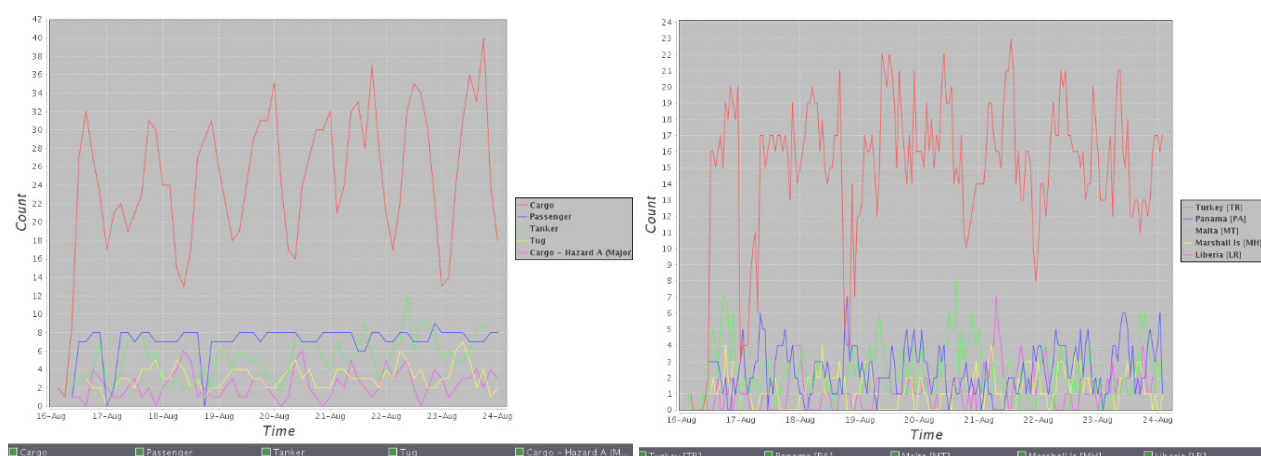


Figure 3-14: VATOZ® Visualizations. Line Charts Showing the Variation of Vessel Type and Flags Along Time (© ASELSAN 2019).

⁶ <https://www.elastic.co/>

Georeferenced data are visualized in a geographical interface using heat and density maps. In Figure 3-15 (top) heat maps represent the density of vessels in the straits, according to their position. The palette varies from red (dense) and blue (sparse). Positions are clustered according to fixed grids, whose color is associated according to the vessel density. Route information, in Figure 3-15 (bottom), is generated as density maps of vessel positions over time.

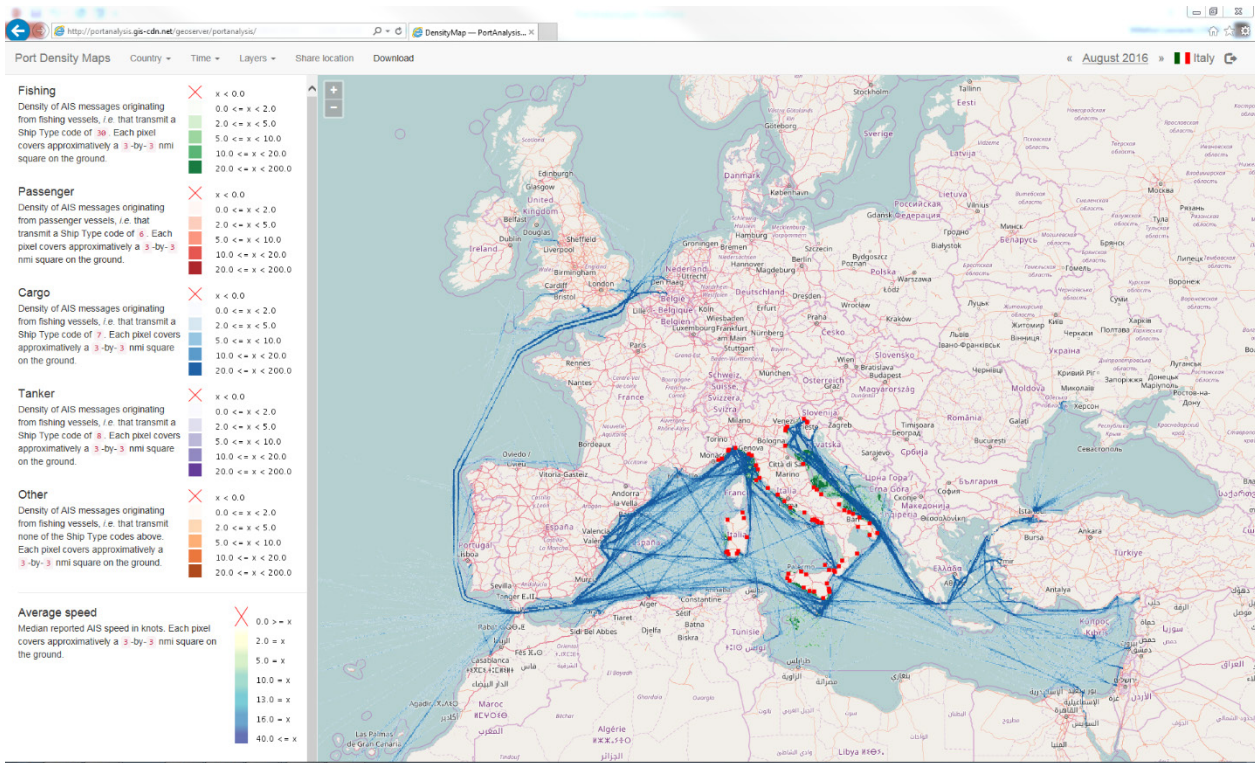


Figure 3-15: VATOZ® Visualizations. Location Based Visualizations for Position and Route Density (©ASELSAN 2019)

3.4.3 Maritime Cyber Security

Geographic views and visual analytic capabilities have been demonstrated as offering a solution to the display of relevant information to the user, amongst the ever-growing amount of data that the systems have to process. As cyber security issues are growing, visualization emerges as a solution. It is particularly important to use visualizations effectively, so that the relevant data are presented to the user in an unambiguous way, reducing false positive information.

Iphar et al. [25] present a visualization interface designed to highlight integrity and veracity issues detected on the AIS. This interface takes as input data the results of threat detection on the AIS messages. The web-based interface enables a dual display showing a map of the maritime traffic but also the list of AIS messages with detected anomalies and associated risks. The interface includes a cartographic layer, few data analytics and a text-based listing of detected features (cf. Figure 3-16 and Figure 3-17). The map relies on two layers: the cartographic layer and the data layer. The cartographic layer constitutes the background of the interface, consisting of Open Street Map (OSM) tiles, enhanced by Open Sea Map (OSeaM) features. The data layer is made of points which are the vessels that have been selected as deserving particular attention by the program. The vessels are shown in different colors according to their vessel type and the user has several options, being able to display all the vessels in the neighborhood of the selected vessel or the elements relative to the vessel itself. The user is also able to discard the vessel if he/she judges that the raising of a

warning does not demonstrate a situation of concern. The corresponding entry in the database is not removed, but is tagged as discarded and is not shown on the screen any longer. This interface aims at offering the people in charge of maritime monitoring a comprehensive overview of the maritime situation in their area of watch.

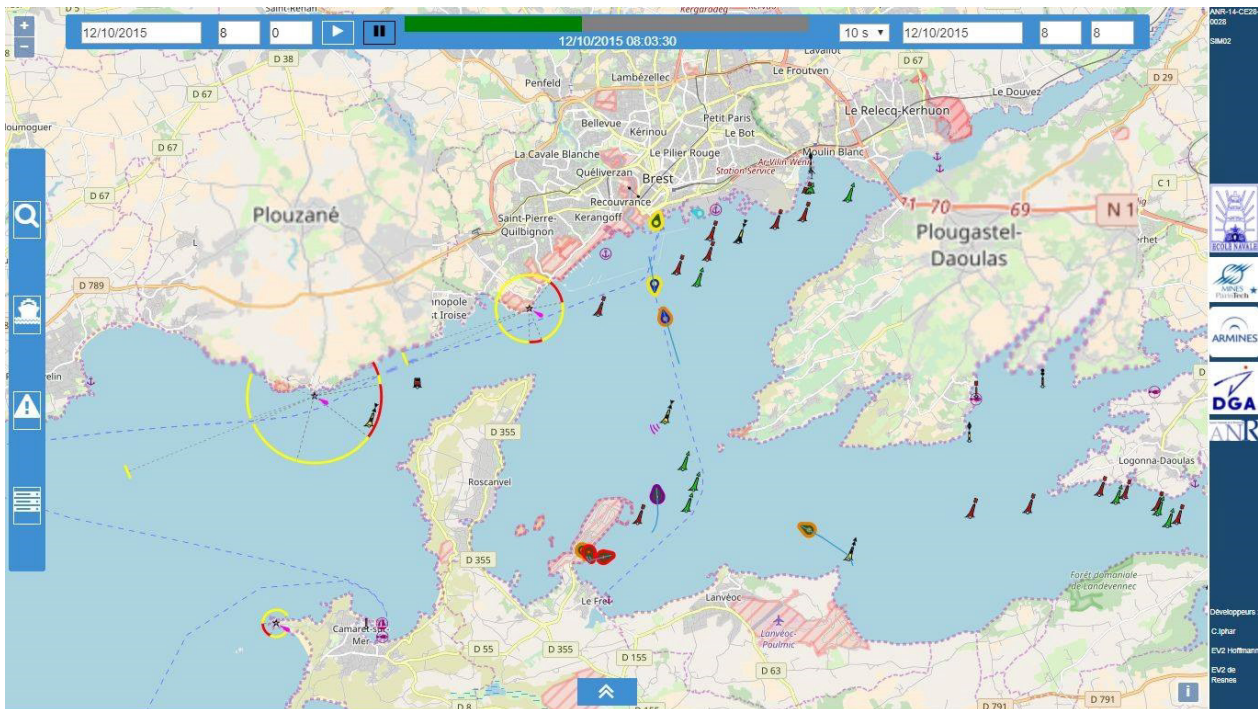


Figure 3-16: Web-Based Interface. Ships are colored depending on the analysis of their AIS messages (© NARI, 2019).

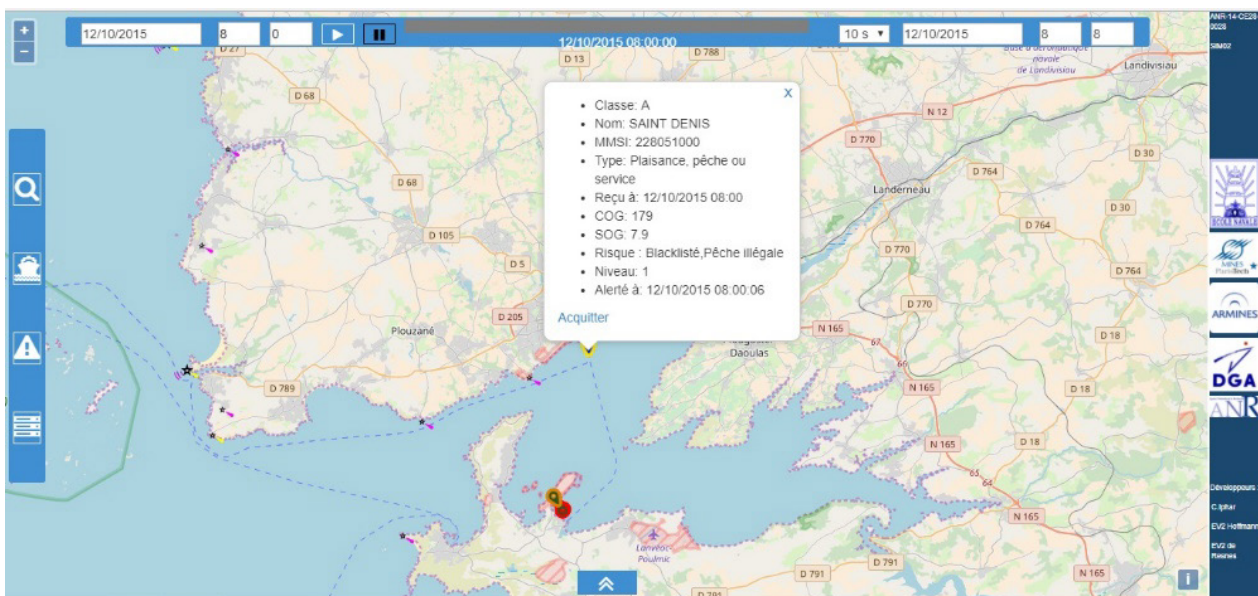


Figure 3-17: Detection and Visualization of an Alert (© NARI, 2019).

3.5 CONCLUSION AND DISCUSSION

This chapter illustrates the work done by IST141/RTG 66 to study innovative interactive visualization and visual analytics for the maritime domain. In order to facilitate the cross fertilization of techniques across domains, maritime datasets, including positional vessel information and contextual data, have been shared within the group and different visualizations and visual analytics software tools developed for research and operational applications have been demonstrated. A survey on interactive visualization and visual analytics developed for maritime applications has been also shared with the group and is included in the chapter.

The maritime domain offers a variety of use cases that can benefit from information visualizations and visual analytics. The literature presents many maritime security situations where the combined use of analytics (e.g., data clustering, predictive analytics), geographical visualizations and aggregated statistics (e.g., on historical data) help the user focusing the attention on relevant events, improving the situation awareness and reducing the user cognitive load. Interactivity is also a fundamental capability of the developed visualizations: the user needs to be able to customize the visualization, as well as to tailor the underlying analytics to the situation.

3.6 REFERENCES

- [1] Ray, C., Dréo, R., Camossi, E., Joussetme, A-L., and Iphar, C. (2019). Heterogeneous Integrated Dataset for Maritime Intelligence, Surveillance, and Reconnaissance. *Data in Brief*, 25(August), 104141 doi: 10.1016/j.dib.2019.104141
- [2] Ray, C., Dréo, R., Camossi, E., and Joussetme, A-L. (2018). Heterogeneous Integrated Dataset for Maritime Intelligence, Surveillance, and Reconnaissance (Version 1.0) [Dataset]. doi: 10.5281/zenodo.1167595
- [3] Vatin, G., and Napoli, A. (2013). High-Level Taxonomy of Geovisual Analytics Tasks for Maritime Surveillance. In *Proceedings of the 26th International Cartographic Conference (ICC 2013)*, Aug 2013, Dresden, Germany, 2013 hal-00856714. <https://hal-mines-paristech.archives-ouvertes.fr/hal-00856714/document>
- [4] Malik, A., Maciejewski, R., Maule, B., and Ebert, D.S. (2011). A Visual Analytics Process for Maritime Resource Allocation and Risk Assessment. *IEEE Symposium on Visual Analytics Science and Technology*, October 23 – 28, Providence, RI, USA, IEEE 2011 doi: 10.1109/VAST.2011.6102460
- [5] Cazzanti, L., Davoli, A., and Millefiori, L.M. (2016). Automated Port Traffic Statistics: From Raw Data to Visualisation. In *2016 IEEE International Conference on Big Data (Big Data)*, Dec 2016, pp. 1569-1573. Doi: 10.1109/BigData.2016.7840765. <https://openlibrary.cmre.nato.int/bitstream/handle/20.500.12489/706/CMRE-PR-2017-010.pdf?sequence=1>
- [6] Andrienko, G., and Andrienko, A. (2011). Exploring Trajectory Attributes in Brest Harbor. *Proceedings of the MOVE COSTS AIS workshop*, Brest, 2011. <http://openaccess.city.ac.uk/id/eprint/2911>
- [7] Last, P., Hering-Bertram, M., Jung, T., and Linsen, L. (2015). Visual Encoding of Automatic Identification System (AIS) Data for Radar Systems. In *Proceedings of the 23rd International Conference in Central Europe on Computer Graphics, Visualisation and Computer Vision (WSCG 2015, short papers proceedings)* Plzen, Czech Republic, June 8 – 12, pp. 49-57 <http://wscg.zcu.cz/WSCG2015/CSRN-2502.pdf>

- [8] Last, P. (2016). Interactive History-Based Vessel Movement Prediction and Visualisation, in Analysis of Automatic Identification System Data for Maritime Safety. Chapter 6, PhD dissertation, Jacobs University Bremen. <https://opus.jacobs-university.de/frontdoor/index/index/docId/690>
- [9] Itani, A., Ray, C., El Falou, A., and Issa, J. (2019). Mining Ship Motions and Patterns of Life for the EU Common Information Sharing Environment (CISE). In Proceedings of MTS/IEEE Oceans, Marseille, France, June 2019. doi: 10.1109/OCEANSE.2019.8867219
- [10] Enguehard, R.A., Hoerber, O., and Devillers, R. (2013). Interactive Exploration of Movement Data: A Case Study of Geovisual Analytics for Fishing Vessel Analysis. Information Visualisation, 12(1), Sage Publications. doi: 10.1177/1473871612456121
- [11] Andrienko, G., and Andrienko, N. (2011). Applying V-Analytics to AIS Data. In Proceedings of the International Workshop on Maritime Anomaly Detection (MAD 2011) June 17 2011, Tilburg, The Netherlands, pp. 25-26. <http://mad.uvt.nl/mad/mad2011-proceedings.pdf>
- [12] Willems, N., van Hage, W.R., de Vries, G., Janssens, G.H., and Malaisé, V. (2010). An Integrated Approach for Visual Analysis of a Multisource Moving Objects Knowledge Base. International Journal of Geographical Information Science, 24(10), pp. 1543-1558. doi: 10.1080/13658816.2010.515029
- [13] Scheepens, R., Willems, N., Van de Wetering, H., Andrienko, G., Andrienko, N., and Van Wijk, J.J. (2011). Composite Density Maps for Multivariate Trajectories. IEEE Transactions on Visualisation and Computer Graphics, 17(12), pp. 2518-252.
- [14] Willems, N., Scheepens, R., van de Wetering, H., and van Wijk, J.J. (2013). Visualisation of Vessel Traffic. In Situation Awareness with Systems of Systems, pp. 73-87. Springer New York. doi: 10.1007/978-1-4614-6230-9_5
- [15] Andrienko, N., Andrienko, G., Camossi, E., Claramunt, C., Cordero Garcia, J.M., Fuchs, G., Hadzagic, M., Jouselme, A-L. Ray, C., Scarlatti, D., and Vouros, G. (2017). Visual Exploration of Movement and Event Data with Interactive Time Masks. Visual Informatics, 1(1), pp. 25-39. doi: 10.1016/j.visinf.2017.01.004.
- [16] Scheepens, R., van de Wetering, H., and van Wijk, J.J. (2014). Contour Based Visualisation of Vessel Movement Predictions, International Journal of Geographical Information Science, 28(5), pp. 891-909. doi: 10.1080/13658816.2013.868466
- [17] Lavigne, V., Gouin, D., and Davenport, M. (2011). Visual Analytics for Maritime Domain Awareness. In Proceedings of the 2011 IEEE International Conference on Technologies for Homeland Security (HST), Waltham, MA, pp. 49-54. doi: 10.1109/THS.2011.6107846
- [18] Lavigne, V. (2014). Interactive Visualisation Applications for Maritime Anomaly Detection and Analysis. In Proceedings of the ACM SIGKDD 2014 Full-day Workshop on Interactive Data Exploration and Analytics (IDEA), pp. 75-84.
- [19] Varga, M.J., and Lavigne, V. (2016). The application of Visual Analytics to Maritime Domain Analysis. IEEE VIS, 23rd – 28th October 2016, Baltimore, USA.
- [20] Riveiro, M., Falkman, G., and Ziemke, T. (2008). Improving Maritime Anomaly Detection and Situation Awareness through Interactive Visualisation. In Proceedings of the 11th International Conference on Information Fusion (FUSION 2008). IEEE Computer Society Press.

- [21] Riveiro, M., and Falkman, G. (2010). Supporting the Analytical Reasoning Process in Maritime Anomaly Detection: Evaluation and Experimental Design. In *Information Visualisation (IV)*, 2010 14th International Conference, pp. 170-178, IEEE Computer Society Press.
- [22] Riveiro, M., Falkman, G., Ziemke, T., and Warston, H. (2009). VISAD: An Interactive and Visual Analytical Tool for the Detection of Behavioral Anomalies in Maritime Traffic Data. In *Proceedings of SPIE*, 46, pp. 07-11.
- [23] Etienne, L., Ray, C., and McArdle, G. (2011). Spatio-Temporal Visualisation of Outliers. In *Proceedings of the International Workshop on Maritime Anomaly Detection (MAD 2011)* June 17 2011, Tilburg, The Netherlands. <http://mad.uvt.nl/mad/mad2011-proceedings.pdf>.
- [24] Ray, C., Goralski, R., Claramunt, C., and Gold, C. (2011). Real-Time 3D Monitoring of Marine Navigation. *Information Fusion and Geographic Information Systems, Towards the Digital Ocean. Lecture Notes in Geoinformation and Cartography*, Springer: Berlin, Heidelberg.
- [25] Iphar, C., Ray, C., and Napoli, A. (2020). Data integrity assessment for maritime anomaly detection, *Expert System with Application*, 147, Elsevier, June 2020, ISSN 0957-4174. doi: 10.1016/j.eswa.2020.113219.



Chapter 4 – EXPLORATORY MEDIA ANALYSIS

Prof. Tomas Krilavičius and Justina Mandravickaitė

Vytautas Magnus University
LITHUANIA

Internet media is one of the most important tools to influence public opinion as well as reflect it. In this report, we analyze reflections of the dynamics of the Ukrainian conflict in BBC, RussiaToday, DayKiev and delfi.lt (main Lithuanian news portal). We apply two different approaches for the analysis: co-occurrence network analysis to reflect the change of language in four different media channels during the conflict as well as sentiment-based storyline (*syuzhet*) analysis to monitor sentiment change in BBC from 2013 to 2014. We split the conflict into three stages: beginning (2013/11/21 – 2014/01/15), escalation (2014/01/16 – 2014/02/17) and occupation of Crimea (2014/02/18 – 2014/02/28). These approaches allow the visual analysis of the conflict dynamics in media. From the application of Artificial Intelligence (AI), Natural Language Processing (NLP) and visualization techniques for big data, we gain a better understanding of the perception of conflict dynamics and public mood on specific topics, and automation of information analysis. Moreover, other types of similar applications are possible [1].

We provide a summary of this research here, see details in Ref. [2].

4.1 DATA AND METHODS

4.1.1 Corpora

We used 2 datasets for our research:

- 1) Results of qualitative discourse analysis of media articles (BBC, RussiaToday, DayKiev and delfi.lt), and
- 2) Raw BBC articles.

The first dataset was prepared by the team of students mentored by scientists during the project “Research Meadow / Mokslo pieva.” Media articles were gathered from 4 different media sources – BBC, RussiaToday, DayKiev and delfi.lt – from 2013 November 21 until 2014 February 28. This period covered 3 stages of the Ukrainian conflict:

- 1) 2013 November 21 – 2014 January 15 (“beginning” stage).
- 2) 2014 January 16 – February 17 (“escalation” stage).
- 3) 2014 February 18 – 28 (“occupation of Crimea” stage).

More details are provided in Ref. [2].

4.1.2 Methods

We have used two methods:

- 1) Word Co-occurrence Networks: they show relations between different media sources based on the co-occurrences of terms (nouns and noun phrases) in their texts.
- 2) Sentiment-Base Storyline analysis: it visually represents sentiment changes in the texts over time.

A detailed description of the methods is provided in Ref. [2].

4.2 RESULTS

In this section, we present the three stages of the conflict as it was covered by four different media sources (BBC, delfi.lt, DayKiev and RussiaToday) that were analyzed using Word Co-occurrence Networks and Sentiment-Based Storyline analysis.

4.2.1 Word Co-Occurrence Networks

Word Co-occurrence networks show dictionary similarities between different media sources, e.g., Figure 4-1 shows that in the 1st stage of the conflict RussiaToday and DayKiev were still using the same terminology, while delfi.lt was in the opposite position, and BBC was representing a number of different opinions. However, in the 3rd stage of the conflict (Figure 4-2) DayKiev rhetoric expression became similar to delfi.lt. For more details and results, see Ref. [2].

4.2.2 Sentiment-Based Storyline Analysis

Sentiment-Based Storyline shows emotion changes in the media sources over time, e.g., Figure 4-3 demonstrates that in the 3rd stage of the conflict, the mood was swinging strongly, based on the different events occurring at the time, namely:

- 1) Clashes begin in earnest: casualties on both sides – protesters as well as police.
- 2) Agreement between opposition and V. Yanukovych (February 21, 2014)
- 3) Y. Tymoshenko is liberated, Yanukovych leaves the country (February 2, 2014).
- 4) February 27 – 28, 2014:
 - 4.1) Russian forces occupy the Crimean parliament building.
 - 4.2) Russian troops are deployed to airports and other strategic sites.

Again, see Ref. [2] for more details and additional examples.

4.3 CONCLUSIONS

The goal of this research was two-fold: first, to evaluate the possibility to analyze media-population interactions during a conflict, and second, to evaluate the possibility to automate the analysis and evaluation processes.

Our research results show that it is comparatively easy to track opinion and rhetoric changes using Natural Language Processing tools. Events in the conflict are well reflected in media, and media partially reflects the change of population perception during the conflict.

While automation was not investigated in detail, in most cases, the analysis was performed with separate tools that could work with minimal human involvement. Of course, several aspects still need to be selected by analysts, namely the conflict itself, media sources (in a full-fledged tool, they just need ticking corresponding checkboxes), and events. We did not investigate automatic or even semi-automatic choice of events; of course, human help can make visualizations better and/or more tailored.

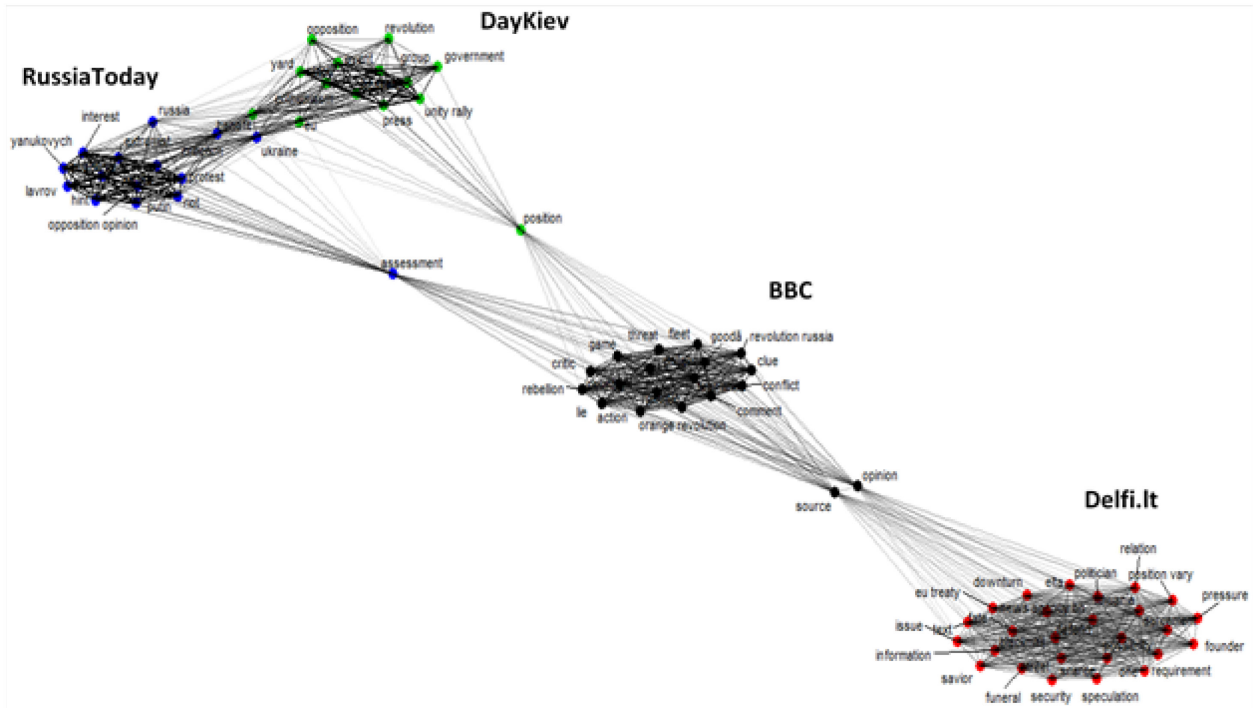


Figure 4-1: Word Co-Occurrence Network: 1st Stage of the Ukrainian Conflict.

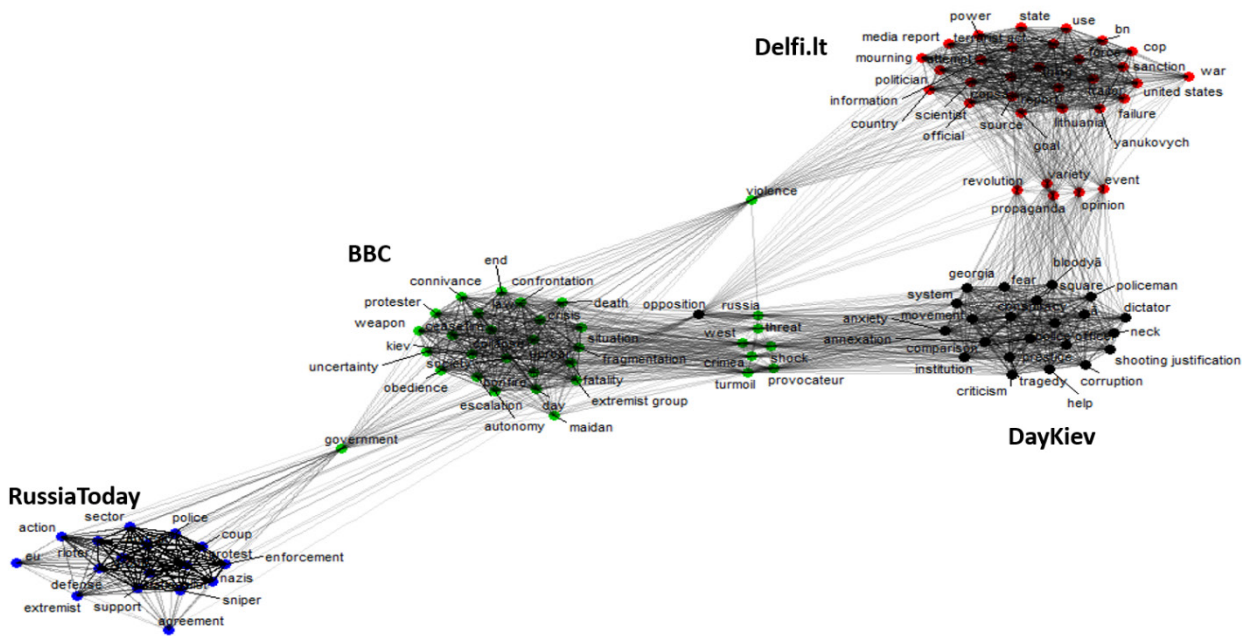


Figure 4-2: Word Co-Occurrence Network: 3rd Stage of the Ukrainian Conflict.

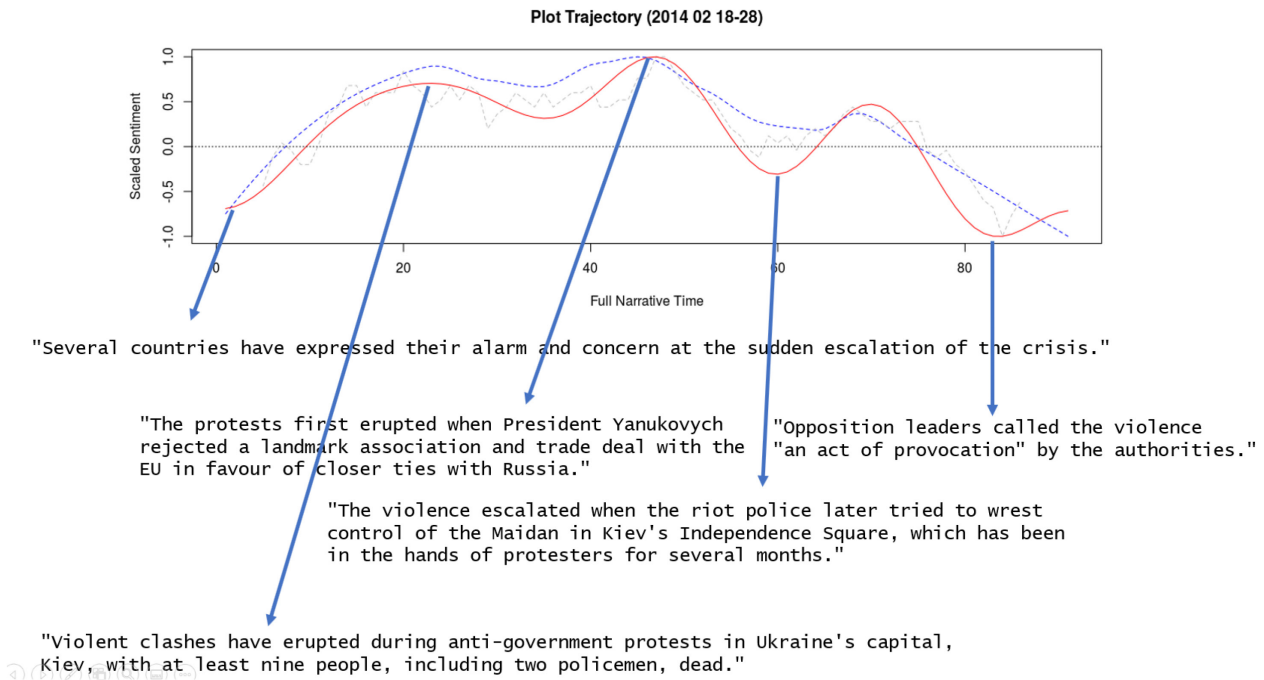


Figure 4-3: Sentiment-Based Narrative Trajectory (with 3 Types of Smoothing: Grey Line – Moving Average, Blue – Loess, Red – *Suyzhet* Discrete Cosine Transformation): 3rd Stage of the Ukrainian Conflict.

Here are the results that were produced in this research:

- 1) Description of two media rhetoric analysis processes: co-occurrence networks and sentiment-based storyline.
- 2) Experimental evaluation of co-occurrence networks approach on Ukrainian conflict reflection in 4 media channels: BBC, RussiaToday, DayKiev and Delfi.lt.
- 3) Experimental evaluation of sentiment-based storyline approach for Ukrainian conflict dynamics in BBC.

Here are our conclusions regarding the capabilities of these visualization tools:

- 1) The Word Co-occurrence Networks approach reflects the change of rhetoric in different media channels. Spatial visualization of networks shows rhetoric differences and similarities between channels.
- 2) The Sentiment-Base Storyline-based analysis reflects a change of opinion over time in a media channel. However, the sentence-wise analysis does not reflect sufficiently well the change over time due to sampling, hence the representation may need a correction regarding distribution patterns.
- 3) Both approaches are suitable for automation.

4.4 REFERENCES

- [1] Mandravickaitė, J., Briedienė, M., Uus, J., and Krilavičius, T. (Feb 2020). What's in the News? Identification of Trending Topics in Alternative and Mainstream Lithuanian Media. MTWD2020, Toulouse, France.

- [2] Mandravickaitė, J., and Krilavičius, T. (2019). Ukrainian Conflict in Media: Two Approaches to Narrative Analysis. Proc. of the Big Data Challenges: Situation Awareness and Decision Support workshop, IST-178, Oct. 15 – 16, 2019, Budapest, Hungary. <https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-178/MP-IST-178-04.pdf>



Chapter 5 – VISUAL EXPLORATION OF SIMULATION DATA

Dr Petter Bivall

Swedish Defence Research Agency
SWEDEN

Simulations are widely used as a safe and often cost-effective means to test systems before production, and as a way to estimate the expected behavior of otherwise inaccessible systems. Common for simulations are chains of events occurring over time, leading to outcomes determined by the simulation's algorithms. This chapter will elaborate on using visualization to analyze simulation conditions and results, as well as on how Visual Analytics (VA) can be applied to better understand the inner life of advanced simulation algorithms.

5.1 INTRODUCTION

Implementing simulation procedures is an efficient way to investigate physical phenomena and chains of events. Applications for simulation techniques are found in a multitude of areas, ranging from simplistic gaming models to, and beyond, Computational Fluid Dynamics (CFD) for simulations of airflow around aircrafts and vehicles, molecular dynamics, and missile tests. The diversity of simulations make them applicable across fields in industry, academia, and military research.

Batch simulation is a simulation methodology where many simulations are run, often in parallel, using slightly different input parameter settings. Examples are variations of drug molecule candidates, weather conditions or the number and formation of tanks used in a battle. The purpose can be to complete the full simulation quickly, or to cover as much of the parameter space as possible in a given time-frame. When the methods include covering a large parameter space, the objective is commonly to search for an optimal solution or, in some cases, to identify the likely properties of unknown or complex systems. As simulation results are generated, there is a need for analysis in order to draw appropriate conclusions from the data. Using the term simulation ensemble to denote what is here referred to as batch simulation, Matković, Gračanin and Hauser [1], state: *“A clever combination of computational and interactive methods supports the simulation expert to gain deeper insight into the data and into the physical phenomenon that is represented by the ensemble.”* In support of using visualization, they also state that *“An analysis environment that combines interactive visualization and computational analysis provides unique advantages for the exploration and analysis of complex ensemble data.”*[1].

The widespread use of simulations in many diverse application areas makes it impossible to present a single method or tool showing the ultimate approach for visualization of all possible kinds of simulation data. Instead, the examples presented here are put forward as a multi-faceted toolbox. The aim is to overview relevant visualization tools and research, both within and outside the simulation field, thereby providing a resource from which visualization techniques can be chosen to best suit the problem at hand. Generalized visualization tools are also presented. Even if they often do not provide the optimal solution, generalized tools are very powerful, have a plethora of representations, and offer an excellent way to start a visual data analysis process.

5.2 VISUAL ANALYTICS FOR SIMULATION DATA

Some applications using batch simulation deliver results with a clear relation to physical domains, such as CFD simulating airflow around aircrafts, or screening for drug molecule candidates to find the one with the strongest binding forces to an enzyme, thereby identifying the optimal inhibitor. Others are more abstract and, although there is a likely relation to the physical world, deliver results such as the chain of events for a simulated hacker attack, or likelihood of survival under different conditions in a combat zone.

Phenomena with an evident mapping to physical properties are often analyzed using methods traditionally labelled as scientific visualization, whereas analysis of more abstract or multivariate data tends to be approached through information visualization tools. Today the overlap between these fields is large, the edges are blurry, and we argue that the most effective visual analytics combine the task-appropriate techniques from both arenas in order to tackle the challenge at hand.

5.2.1 Post Analysis

Visual post analysis is still the most common way to investigate simulation data. As the name suggests, post analysis is performed after the simulation has completed, preferably using a mix of visualization techniques to enable a detailed exploration of the data.

The work of Benjaminsson et al. [2] is one example of how simulation data can be successfully presented using hybrid visualization techniques with both abstraction and connections to a physical context. The data used in Ref. [2] was from a large-scale brain simulation, and their first visualization approach falls into the category of post analysis. Using animated glyphs of varying sizes and colors they abstracted signaling between neurons in the brain, thus representing both the amounts of signaling and when it occurred. Also, optional placing of the data representation in the context of a wireframe model of a brain provided an informative view of which part of the brain the simulation covered.

The work by Li et al. [3] focuses on analysis of simulation data in the area of Earth System Sciences (ESS). They present a new version of a system named VAPOR, which focuses on being able to work with large amounts of geo-referenced data; Li et al. point out the lack of support for such data in many generalized visualization tools. Examples of their contributions include functions for handling multi-resolution datasets to enable use of VAPOR on commodity hardware, easy access to tools for ESS data treatment otherwise only found in command-line software, and stable handling of missing data points.

Fang and Zhang [4] look specifically at behavioral trajectory pattern recognition in longitudinal data with a large number of parameters. This might seem far from simulation data; however, finding a visualization technique tailored for your own specific needs out-of-the-box is unfortunately unlikely, thus a willingness to look into technology transfer is required. The visual representation and data processing presented in Ref. [4] might not be applicable for a multitude of simulation data but can prove powerful if the simulation is based on a large number of parameters and results are likely to include clusters.

Using a collection of common and established representations for visual analytics can be very powerful if the representations are chosen carefully with respect to the application area. The work presented in Refs. [5], [6] does not introduce novel visual representations; however, the work exhibits a very good example on how interconnecting visual representations of data can aid in producing new findings, and how development of a well-constructed composite also is a good contribution to the visualization community. The study presented in Ref. [5] uses visual analytics techniques to study performance of differently configured simulations in a fairly complex simulation design. In their parallel discrete-event simulation, operations are executed in a somewhat opportunistic manner, which means not necessarily in the correct order with respect to the timeline. When operations are verified against the simulation's virtual time, the design requires some operations to be rolled back and re-executed under somewhat different conditions. These rollbacks come with a performance cost. By gathering data during simulations and developing a visual analysis tool for this specific application, Ross et al. could successfully analyze information flows, compare the performance costs associated with the rollbacks, and how frequently rollbacks occurred when using different configurations [5].

Knowledge discovery and easing understanding for non-experts are important parts of the work by Feldkamp, Bergman and Strassburger [7]. Like the work in Ref. [5], [6] they implement well-known visual representations such as Parallel Coordinates Plots (PCP) and scatter plots to produce an interactive display of

the simulation parameters and output. The primary aim for the work in Ref. [7] is to facilitate insights about the simulated object or process, and they suggest using combinations of visualization and machine learning techniques.

Three important contributions in the area of visualization are The Visualization Tool Kit (VTK) [8], ParaView [9] and VisIt [10]. They are research and development efforts in their own right, and there are also many examples of research where these tools have been implemented and extended. When viewed together they form a respectable ensemble and a nice example of how research efforts can build on top of each other: VTK is a software library designed for visualization, including data readers, data filters and renderers. ParaView brings VTK into an extended Graphical User Interface (GUI), providing the user with access to most VTK functions without programming. VisIt also builds on VTK (among other libraries) and can be used for post analysis of very large datasets. Furthermore, together with its libsim library, VisIt expands into real-time visualization of ongoing simulations. These tools are all open source, and further described in Section 5.3.

5.2.2 In situ Analysis

Using visualization tools for monitoring simulations and computational steering is not a novel concept, and the examples in Refs. [11], [12], [13], [14] are just a few, with the oldest dating back over twenty years. At the time of writing, when studying the corpus of visualization and simulation research, in situ visualization rises as the currently most popular field in visualization for simulations. Here in situ refers to visualization presenting the results from the simulation simultaneously as the simulation is running, sometimes also allowing for computational steering through live adjustments of the parameters regulating the simulation. The approach of in situ visualization has gained ground as increased computer power has allowed for running more simulations, or simulations with higher fidelity, both cases generating larger amounts of data to transfer and store. Part of the idea with in situ visualization is to avoid the bottleneck of data transfer by direct observation of the simulations' results Refs. [2], [15], [16].

The post analysis approach of Benjaminsson et al. is described in Section 5.2.1. In their second approach, they implemented in situ visualization providing a continuous view of the simulation. For this purpose, they used the open source software VisIt [10].

A more recent example of in situ visualization is the work by Buffat et al. [17]. They present creative use of ParaView and VisIt, with the addition of Python libraries, together comprising a powerful suite for online simulation supervision and steering. With a High Performance Computing (HPC) system solving Navier-Stokes equations, Buffat et al. investigate different ways to analyze and interact with the simulation. A hybrid approach is concluded to provide the best performance, using a high number of computational nodes working on the simulation, and fewer nodes dedicated to computational steering and monitoring the simulation as it develops over time.

Virtual environments with simulations representing real-world events can be considered a form of in situ analysis, as events in the simulations produce immediate observable visual feedback in the virtual environment. Bijl and Boer [18] look into using game technology for visualization during simulations. They conclude that when working with training environments, improving realism in the visual representations is advantageous.

Whereas most examples of in situ analysis work with visual representations of the simulated objects or phenomena, the work by Brunhart-Lupo et al. [19] present a visual analysis and simulation control system working in parameter and output value spaces. The approach taken is to use parallel planes in an immersive environment. The parallel planes are similar to a parallel coordinates plot, although where PCP displays one axis for each data column, or parameter, each parallel plane renders an intersection point for two parameters, see Figure 5-1.

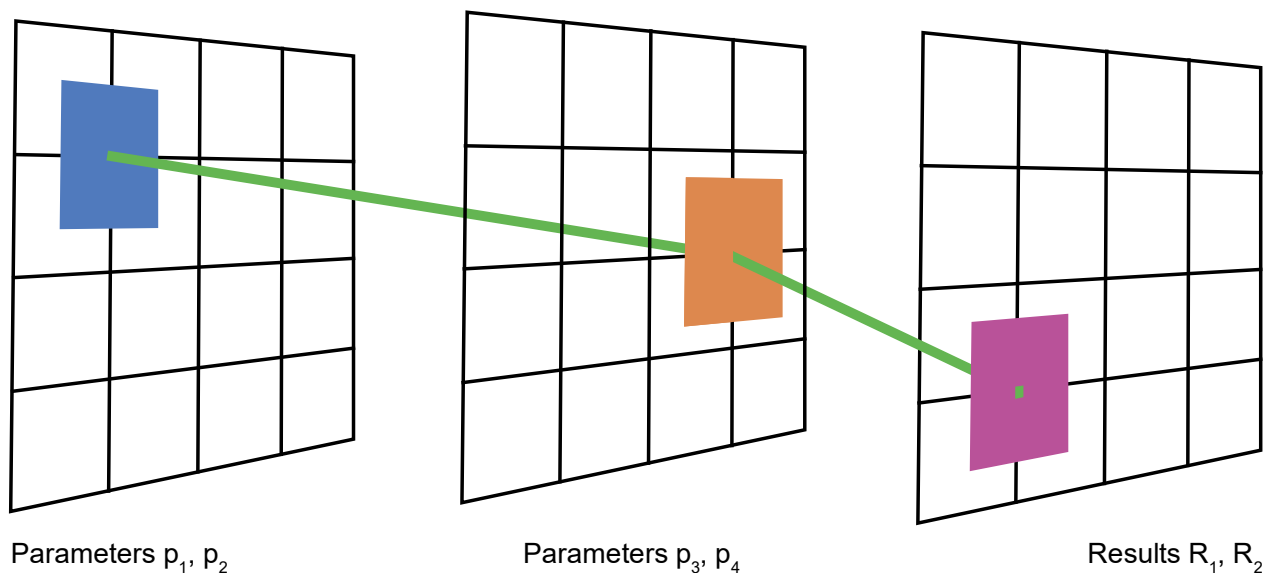


Figure 5-1: Basic Principle of a Plane-Based PCP. Each line intersects the plane at the point corresponding to the “coordinate” formed by the values of the two parameters used for the plane. Selections for data filtering can be applied to each plane.

The work in Ref. [19] shows an interesting approach to the coupling between simulation and visualization, using a hybrid between post analysis and parameter space in situ visualization. Data resulting from the simulations are rendered in the visual representation and, at the same time, selection (or brushing) in the parallel planes showing input parameter data triggers simulations using the selected portions of the parameter space. Thereby the method provides control of the simulation parameters and, as soon as the simulations are complete, rapid visual feedback of the output data.

In a manner similar to Ref. [19], but using more traditional 2D tools, Matković et al. [1] also apply visual analytics with abstract visual representations for both analysis of simulations results and as the means to selecting parameter ranges for new simulations. Matković et al. promotes the combination of interactive visual analysis and computational methods, and also emphasize the importance of the visualization tool design, for example regarding coupled views with linked selection (brushing), and consistency throughout the user interface.

5.3 SOFTWARE TOOLS SUMMARY

One of the aims of the collaborative work in IST-141/RTG066 was to share experiences and tools for visual analytics between the members of the group. This section will describe a selection of tools applicable in the area of analyzing data from batch simulations, aiming at understanding the outcomes as well as the simulation processes.

5.3.1 Time Line Graphs

Time Line Graphs (TLG) is a VA tool for analysis of multi-channel time-dependent data. The tool, being developed at the Swedish Defence Research Agency, presents the user with a set of interactive line graph matrices, where each matrix shows time synchronized views of data originating from one or multiple sources, see Figure 5-2.

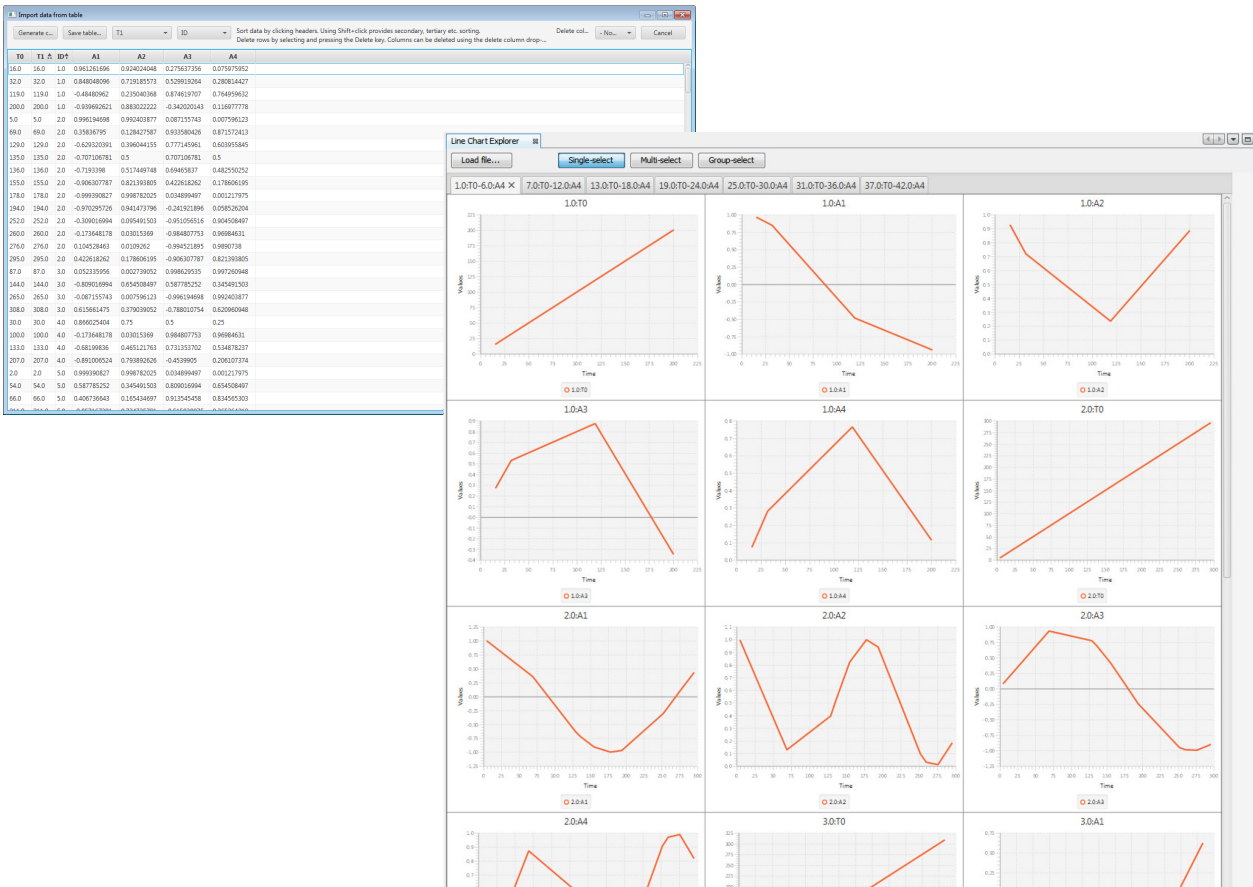


Figure 5-2: Time Line Graphs Example, Showing the Table Data Importer (Top) and the Multi-Tabbed Graph View (Bottom).

Using a single representational form imposes some restrictions on the use of TLG, as compared to using multiple representations with each potentially highlighting different features in the data. However, for TLG the main purpose is to enable analysis of data with a common time-frame. One salient example generating such data is simulations of dynamic events.

Generally, a simulation model continuously updates many internal states, and events are triggered based on certain criteria to generate discrete state changes in the model. Some models are also too complex to overview easily, such as some simulations based on genetic algorithms where abundant data is generated and processed. If simulation output data are logged together with internal states and signals from the simulation model, it becomes possible to analyze how input, output and internal states of the simulation models vary over time, allowing for a better understanding of the simulation and its results.

For this purpose, a lot of effort has been put into empowering the user with possibilities to interact with the data representations, implementing overviews, zooming, and filtering, and extra details on demand [20]. Initially graphs are generated into a matrix/grid structure and placed onto tabs, with each grid containing as many graphs as the user requests. Graphs are also optionally grouped according to the structure of the data table if a group identifier is provided. Grouping can be used to indicate, for example, which part of a simulation model the data originates from.

Once the initial state of the application is set, with a data table translated into grids of charts, the user's analysis process begins. This can be a targeted or explorative analysis depending on the application. The multiple-charts overview provides an opportunity to search for interesting features such as discrete events or

chronologically dependent changes in parameter values, e.g., occurring at the same or sequentially. Charts can be sorted by the user and placed in a common tab/grid for further analysis. Charts can also be grouped to highlight common factors, a useful feature when comparing multiple charts in the analysis process. Further features include, among others, the ability to drill down into charts by zooming, selecting for details on data values, and a data filter structure that can be applied to each chart in order to extract derived data such as the rate of change or clusters.

With the human interaction required, the number of charts cannot be too extensive. Thus, the TLG tool can be applicable for large datasets, but not for extensive Big Data analysis.

5.3.2 The Visualization Toolkit and ParaView

The Visualization Toolkit (VTK) [8] is a mature software framework for the complete visualization pipeline. If your goal is to produce a custom tool for visualization of your scientific data, VTK provides a lot of functionalities out-of-the-box, and well-structured interfaces if extensions are required. With a small amount of code it is possible to read many different data formats, apply several kinds of data filters, and map data to visual representations rendered on screen.

Although you can build GUIs with the components offered by VTK, this is not the framework's strong suit. Interaction models are provided, and it is fairly easy to set up a window for visualization, but whereas data treatment and mapping seems efficient, the coding efforts needed to produce a VTK-based GUI is probably better spent on other features. If the aim is to rapidly get an interactive rendering of your data, ParaView is a better choice.

ParaView [9] is a powerful application for data visualization with capability to do distributed rendering of very large datasets. Here the focus will be on using ParaView as a desktop visualization tool, leaving functions for distributed rendering to further reading. ParaView is actually based on VTK and offers access to most of the functions you find in VTK, but without the need for coding. The learning curve for using ParaView is steep, but tutorials are available. The documentation is good and ParaView has a large and active user community.

If you have worked with VTK the workflow in ParaView will be familiar. First, data is loaded with one of the many data readers, and then data is treated using filters. Filters can be applied for trimming data to a subset, forming vectors from multiple columns in a table, or any other operation offered by the multitude of filter functions available. As operations are added onto others a tree is formed, this can be seen in the upper left part of Figure 5-3 and Figure 5-4. The GUI assists by indicating which filters you can apply to the current data as nodes in the tree are selected.

The pictures in Figure 5-3 and Figure 5-4 show two different types of data and examples of visual representations in ParaView. Figure 5-3 shows a dataset from a simulation of a heated and spinning disc, delivered with ParaView, and used in its tutorials. The dataset contains multiple types of data such as scalars for temperature and vectors for velocity. Figure 5-3 shows visual representations commonly used for these kinds of data, such as streamlines, glyphs and coloring schemes based on data values.

A different example is presented in Figure 5-4, which shows a computed tomography (CT) dataset. The figure further displays ParaView's capability to render multiple views simultaneously. The left view is a volume rendering of the data, using a separate coloring and opacity scheme for a subset of the data, seen as a blue section of the spine. The top right view shows a single slice of the volume, using x-ray like coloring, and the lower right is a histogram presenting the distribution of data values in the volume. The rightmost part of the figure shows the editor for mapping data values to coloring and opacity, known as a transfer function.

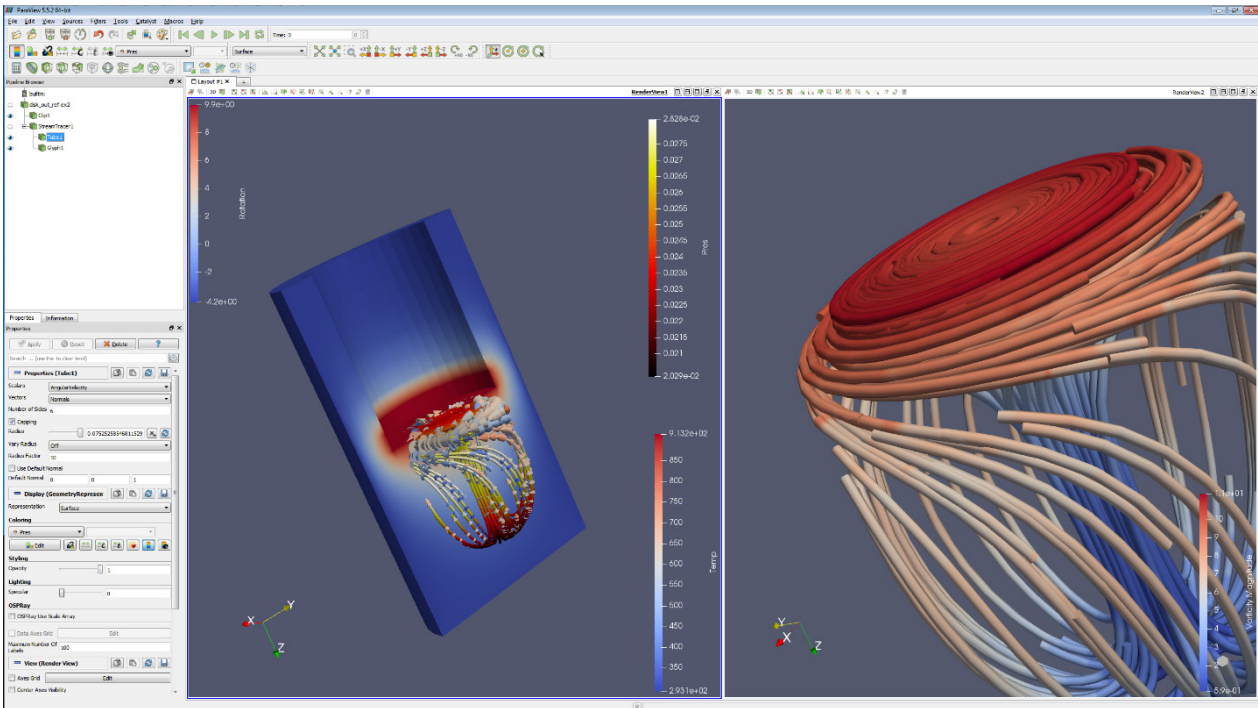


Figure 5-3: ParaView Example Showing a Split View, Where Each View Can be Used to Emphasize Different Features in the Data, Both by Shape and Color Schemes. The data is an example delivered with ParaView, consisting of simulation output describing the airflow around a heated and spinning disk.

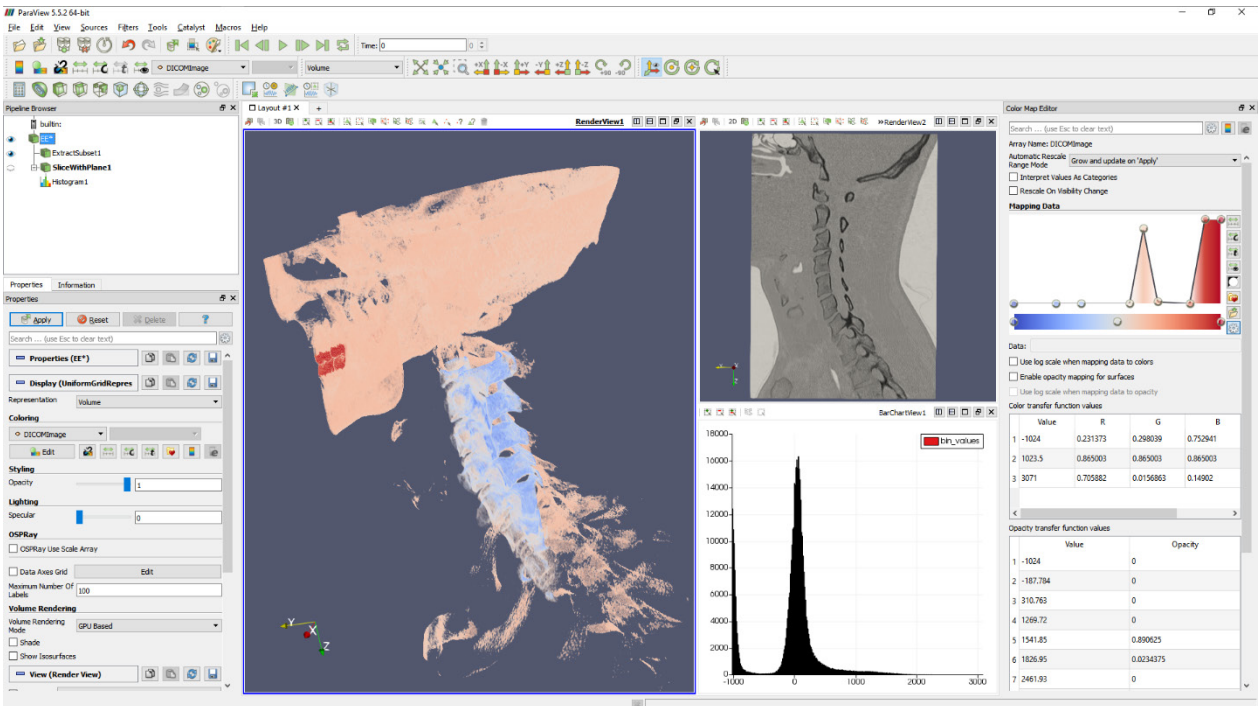


Figure 5-4: ParaView Example Showing CT Data in a Volume Rendering (Left), an X-Ray Like Slice View (Top Right), a Data Histogram (Bottom Right) and a Transfer Function Editor (Rightmost Column).

5.3.3 VisIt

Temporal/chronological aspects and the inner workings of simulations were of special interest for this chapter. The software VisIt is relevant in this context as it has the potential to provide a window into the state of a simulation and showing data in a time-dependent manner, possibly giving further insights into how, why, and when events occur in a simulation. VisIt is a mature open source software that can be used for post analysis but additionally allows for in situ visualization as it includes parts for interfacing and controlling simulations. VisIt uses components from the ParaView application for visualization, see Ref. [10] for details.

VisIt is frequently used in the scientific simulation community, especially when applying in situ visualization. In the overview of current in situ visualization research presented in Section 5.2.2, there are several examples where VisIt has been used or is mentioned.

5.4 CONCLUSIONS AND DISCUSSION

In this chapter, the main categorization for visualization tools aimed at simulation data was post analysis and in situ visualization. Another way to skin the simulated cat is dividing into concrete visual representations showing the real-world events or phenomena, e.g., Refs. [2], [3], and abstract visual representations showing data without further simulation context, e.g., Refs. [1], [5], [6]. Regardless of categorization, in some cases the approach is to trigger new simulations immediately on interaction with the visual representations, and in some cases the tools are used to analyze data and configure new simulation runs for further analysis. The findings from the present overview of the field shows many good examples of merging what has previously been divided into scientific visualization and information visualization.

There is a plethora of different tools for visual analysis of simulation data; however, applicability of existing tools are highly dependent on the data being analyzed and its context. As the authors of Ref. [3] point out, general purpose visualization packages are not always easy to use efficiently for domain specific tasks. Without resources to spend on customizing software, the best way forward can be to assemble a toolbox of different software, together covering the visualization needs. Sometimes the search for suitable tools must extend outside your own field. The work by Fang and Zhang [4] is one such example, having a rather specific intended application, but affording techniques that could be applicable under the right simulation conditions.

Geo-located data is often a crucial component in a military context, making VAPOR [3] a candidate for technology transfer as the tool supports geo-referenced data. Although on a different scale, the approach of showing data in a surrounding context also ties in well with the design chosen in Ref. [2]. There the data was not only presented on an abstract canvas for analysis but rendered combined with a 3D object providing the data with a framing in a physical context.

Despite not being optimally designed for any particular visualization task, the generalized tools presented in this chapter offer very good place to start. Using the wide range of visual representations provided the tools can support proof of concept for many kinds of applications for visualization, without writing a single line of code. It is likely you will find new ways to view your simulation output once you start your visual explorations.

5.5 REFERENCES

- [1] Matković, K., Gračanin D., and Hauser, H. (2018). Visual Analytics for Simulation Ensembles. In 2018 Winter Simulation Conference (WSC).

- [2] Benjaminsson, S., Silverstein, D., Herman, P., Melis, P., Slavnić, V., Spasojević, M., Alexiev, K., and Lansner, A. (2012). Visualization of Output from Large-Scale Brain Simulations. Partnership for Advanced Computing in Europe (PRACE), Project ID: PRPC06.
- [3] Li, S., Jaroszynski, S., Pearse, S., Orf, L., and Clyne, J. (2019). VAPOR: A Visualization Package Tailored to Analyze Simulation Data in Earth System Science. *Atmosphere*, 10.
- [4] Fang, H., and Zhang, Z. (2018). An Enhanced Visualization Method to Aid Behavioral Trajectory Pattern Recognition Infrastructure for Big Longitudinal Data. In *IEEE Transactions on Big Data*, 4(2), pp. 289-298.
- [5] Ross, C., Carothers, C.D., Mubarak, M., Carns, P., Ross, R., Li, J.K., and Ma, K. (2016). Visual Data-Analytics of Large-Scale Parallel Discrete-Event Simulations. In *2016 7th International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS)*.
- [6] Li, J.K., Fujiwara, T., Kesavan, S.P., Ross, C., Mubarak, M., Carothers, C.D., Ross, R.B., and Ma, K. (2019). A Visual Analytics Framework for Analyzing Parallel and Distributed Computing Applications. In *2019 IEEE Visualization in Data Science (VDS)*.
- [7] Feldkamp, N., Bergmann, S., and Strassburger, S. (2015). Visual Analytics of Manufacturing Simulation Data. In *2015 Winter Simulation Conference (WSC)*.
- [8] Schroeder, W., Martin, K., and Lorensen, B. (Eds.) (2006). *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*, 4th Ed. Clifton Park, N.Y: Kitware.
- [9] Ahrens, J., Geveci, B., and Law, C. (2005). ParaView: An End-User Tool for Large-Data Visualization. In C.D. Hansen and C.R. Johnson, Eds., *Visualization Handbook*, Burlington, Butterworth-Heinemann, pp. 717-731.
- [10] Childs, H., Brugger, E., Whitlock, B., Meredith, J., Ahern, S., Pugmire, D., Biagas, K., Miller, M., Harrison, C., Weber, G.H., Krishnan, H., Fogal, T., Sanderson, A., Garth, C., Bethel, E.W., Camp, D., Rübel, O., Durant, M., Favre J.M., and Navrátil, P. (2012). VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data. In *High Performance Visualization—Enabling Extreme-Scale Scientific Insight*, Chapman and Hall/CRC, pp. 357-372.
- [11] Kerlick, G.D., and Kirby, E. (1993). Towards Interactive Steering, Visualization and Animation of Unsteady Finite Element Simulations. In *Proceedings Visualization '93*.
- [12] Johnson, C., Parker, S.G., C. Hansen, Kindlmann, G.L., and Livnat, Y. (1999). Interactive Simulation and Visualization. *Computer*, 32(12), pp. 59-65.
- [13] Watanabe, T., Kume, E., and Kato, K. (2002). Simulation Monitoring System Using AVS. In *Computational Science – ICCS 2002*, Berlin.
- [14] Larsen, M., Brugger, E., Childs, H., Eliot, J., Griffin, K., and Harrison, C. (2015). Strawman: A Batch in situ Visualization and Analysis Infrastructure for Multi-Physics Simulation Codes. In *Proceedings of the First Workshop on in situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization*, New York, NY, USA.
- [15] Rivi, M., Calori, L., Muscianisi, G., and Slavnic, V. (2012). In-Situ Visualization: State-of-the-art and Some Use Cases. PRACE White Paper.

- [16] Valenta, P. (3 Aug 2016). In Situ Visualization Technique. [Online]. <https://summerofhpc.prace-ri.eu/in-situ-visualization-technique/> Accessed 14 Oct 2019.
- [17] Buffat, M., Cadiou, A., Penven L., and Pera, C. (2017). In Situ Analysis and Visualization of Massively Parallel Computations. *The International Journal of High Performance Computing Applications*, 31, pp. 83-90.
- [18] Bijl, J.L., and Boer, C.A. (2011). Advanced 3D Visualization for Simulation Using Game Technology. In *Proceedings of the 2011 Winter Simulation Conference (WSC)*.
- [19] Brunhart-Lupo, N., Bush, B.W., Gruchalla, K., and Smith, S. (2016). Simulation Exploration Through Immersive Parallel Planes. In *2016 Workshop on Immersive Analytics (IA)*.
- [20] Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*.

Chapter 6 – EXPLORING DEEP LEARNING

Valérie Lavigne

Defence Research and Development
Canada

6.1 INTERACTIVE VISUALIZATION AND DEEP LEARNING RESEARCH

Recently, Deep Learning (DL) has had an important impact on a variety of Artificial Intelligence (AI) problems. However, as we increasingly rely on machine learning, there is a need for humans to be in a position to interpret the internal operations of deep neural networks and to explain the models' outputs. Visual Analytics (VA) can contribute to the understanding and leveraging of AI for a number of defence and security tasks.

A short literature review was performed to explore the use of VA in conjunction with DL. Looking up the terms visualization, VA, DL, and deep neural networks, we gathered research papers on this topic. From our analysis of these papers, we concluded that VA techniques can be applied to enhance DL approaches related to three goals:

- 1) Understanding deep neural networks internal operations;
- 2) Explaining DL results; and
- 3) Exploiting the synergy between VA and DL.

Comprehensive reviews of VA applied to understanding and explaining DL were recently published [1], [2], [3], and we leveraged those studies to list many of the VA approaches applied to the first two goals above (see Figure 6-1). We also considered a comprehensive state of the art study on integrating machine learning into VA [4] and a study which focused more specifically on predictive VA [5], although these studies did not discuss DL specifically, they did summarize the advances made at the intersection of machine learning and VA. Beyond explaining AI results, VA approaches which relate to applications leveraging the synergy between VA and DL are very promising.

6.2 UNDERSTANDING DEEP NEURAL NETWORKS INTERNAL OPERATIONS

VA is emerging as a promising research field for enabling humans to better understand the inner workings of deep neural networks. This type of visualization informs researchers to help develop better architectures, better training processes, and ultimately better applications. VA's rich user interactions can indeed provide insights for model debugging and refinement.

Some interactive visualization tools are intended for educational purposes, to make machine learning more accessible, as explained by Choo and Liu:

Tensorflow Playground [6] is an effective system for education and intuitive understanding, where users can play with simple neural networks by changing various configurations in terms of the numbers of layers and nodes, and the types of nonlinear units. It adopts two-dimensional toy data sets for classification and regression tasks. The manner in which each node in the network is activated across different input data values is fully visualized as a heatmap in a two-dimensional space. Another web-based deep learning library called ConvNetJS [7] features easy access to deep learning techniques by simply using a web browser, together with a rich set of visualization modules, which can be effectively used for education and understanding [3].

Visual Analytics in Deep Learning

Interrogative Survey Overview

§4 WHY

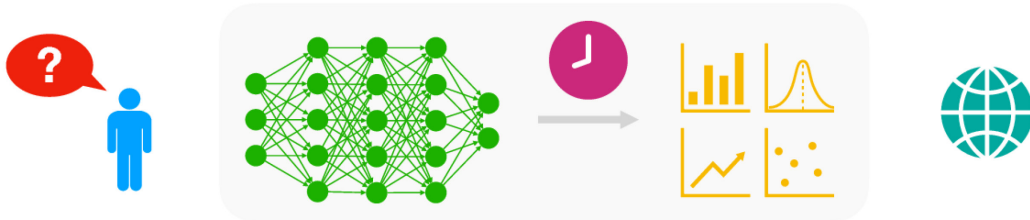
Why would one want to use visualization in deep learning?
 Interpretability & Explainability
 Debugging & Improving Models
 Comparing & Selection Models
 Visualization for Bias Detection
 Teaching Deep Learning Concepts

§6 WHAT

What data, features, and relationships in deep learning can be visualized?
 Computational Graph & Network Architecture
 Learned Model Parameters
 Individual Computational Units
 Neurons In High-dimensional Space
 Aggregated Information

§8 WHEN

When in the deep learning process is visualization used?
 During Training
 After Training



§5 WHO

Who would use and benefit from visualizing deep learning?
 Model Developers & Builders
 Model Users
 Non-experts

§7 HOW

How can we visualize deep learning data, features, and relationships?
 Node-link Diagrams for Network Architecture
 Dimensionality Reduction & Scatter Plots
 Line Charts for Temporal Metrics
 Instance-based Analysis & Exploration
 Interactive Experimentation
 Algorithms for Generating Synthetic Images

§9 WHERE

Where has deep learning visualization been used?
 Application Domains & Models
 A Vibrant Research Community

Figure 6-1: A Visual Overview of Interrogative Questions About VA in DL. Reproduced with permission from the author. The original can be found in Ref. [1].

Most DL frameworks also come with their own visualization tools for low-level monitoring of network training and inference, such as TensorBoard [8], [9] (for TensorFlow), Visdom [10] (for PyTorch), and DL4J Training UI [11] (Deep Learning for Java). VA can also contribute to model training by allowing the user to label data which indirectly integrates user inputs into the model construction.

Many VA tools have been developed for deeper neural network model diagnosis and understanding, particularly for computer vision, Natural Language Processing (NLP), and time series analysis applications.

The technique developed by Zeiler and Fergus [12] inspired many researchers using Convolutional Neural Network (CNN) models in computer vision. CNNVis [13] is a VA tool using deconvolutions to enable a better understanding of the object classification process by visualizing the activations produced by each layer of a CNN as it processes an image or a video (see Figure 6-2).

ReVACNN [14] is another interactive visualization for CNNs which allows exploring and steering the network by visualizing its layers and nodes. Additionally, it provides a filter-level 2D embedding view [15] by applying t-SNE [16] to various filter information, such as filter coefficients, filter gradients, the activation maps, and the activation gradients.

The two previous VA tools were specific to CNN models. RNNVis [17] presents a VA method for understanding and comparing Recurrent Neural Network (RNN) models for NLP tasks (see Figure 6-3). It explains the function of individual hidden state units based on their expected response to input texts. Long Short-Term Memory (LSTM) is an artificial Recurrent Neural Network (RNN) architecture with feedback

connections for sequence processing. LSTMVis [18] is another visual analysis tool for RNN with a focus on understanding the hidden state dynamics behind sequence processing (see Figure 6-4).

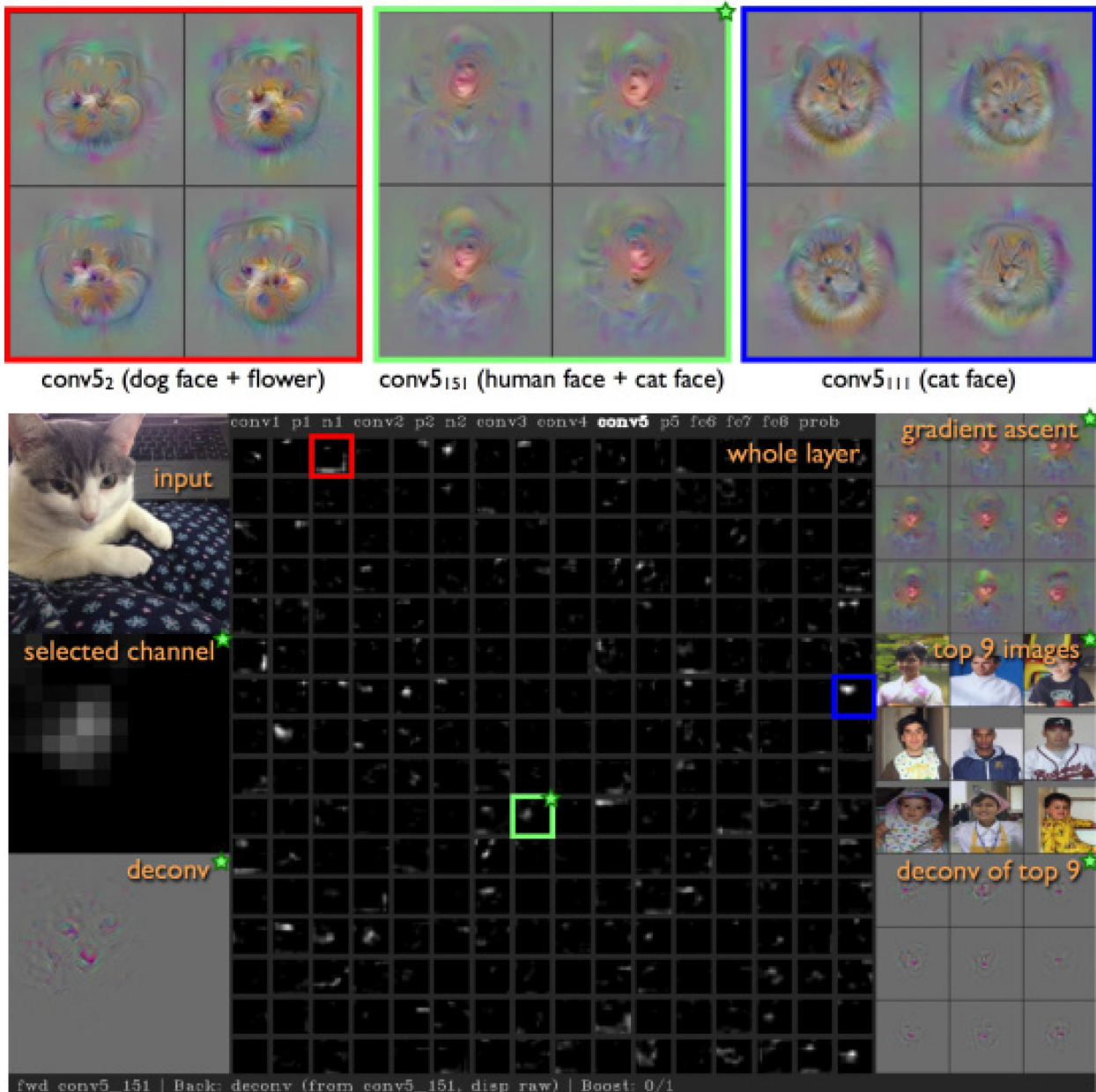


Figure 6-2: CNNVis: The Bottom Shows a Screenshot from the Interactive Visualization Software. The webcam input is shown, along with the whole layer of conv5 activations. The selected channel pane shows an enlarged version of the 13 x 13 conv5₁₅₁ channel activations. Below it, the deconvolution starting at the selected channel is shown. On the right, three selections of nine images are shown: synthetic images produced using the regularized gradient ascent methods, the top 9 image patches from the training set (the images from the training set that caused the highest activations for the selected channel), and the deconvolution of the top 9 images. Reproduced with permission from the author. The original can be found in Ref. [13].

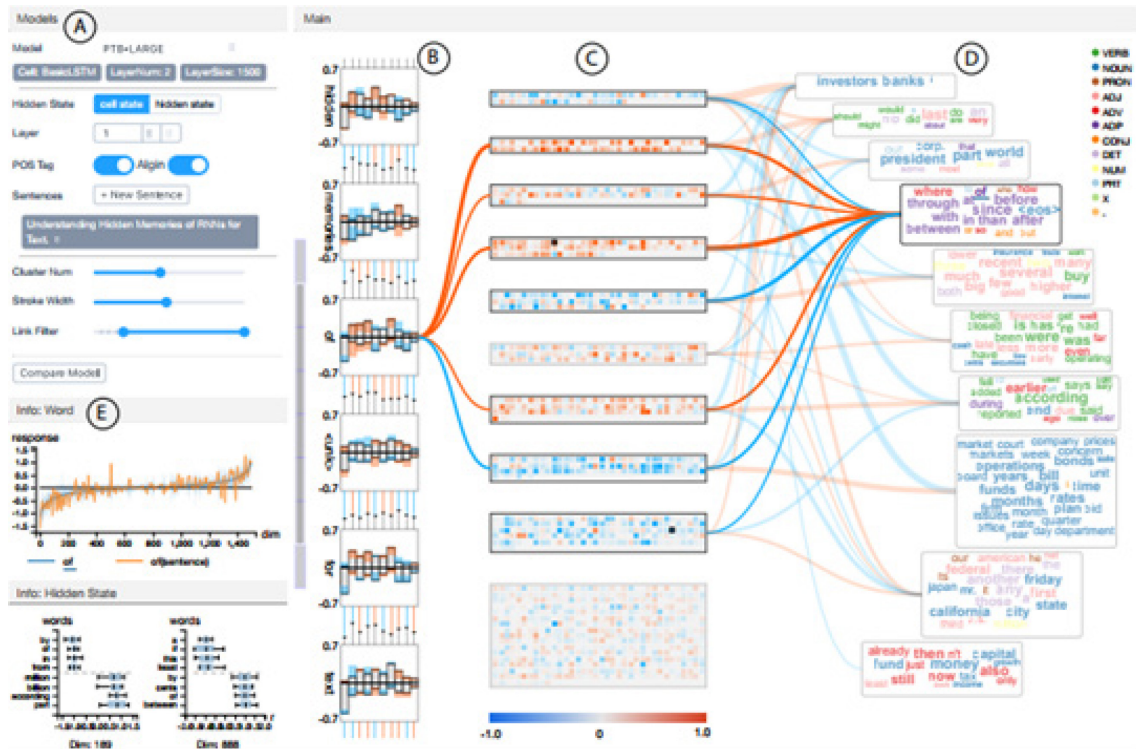


Figure 6-3: The Interface of RNNVis. The control panel (A) shows parameter settings of an RNN and allows users to adjust visualization style. The main view (B-D) contains glyph-based sentence visualization (B), memory chips visualization for hidden state clusters (C), and word clouds visualization for word clusters (D). The detail view (E) shows the distributions of models' responses to selected word "of" and interpretations of selected hidden units. Reproduced with permission from the author. The original can be found in Ref. [17].



Figure 6-4: LSTMVis User Interface. The user interactively selects a range of text specifying a hypothesis about the model which is then used to match similar hidden state patterns. Reproduced with permission from the author. The original can be found in Ref. [18].

ActiVis [19] was also created for local inspection of neuron activations but can handle complex models and large datasets. It provides integrated multiple coordinated views (see Figure 6-5), including a computation graph overview of the model architecture, and a neuron activation view for pattern discovery and comparison, where users can explore complex deep neural network models at both the instance- and subset-level. “ActiVis visualizes how neurons are activated by user-specified instances or instance subsets, to help users understand how a model derives its predictions. Users can freely define subsets with raw data attributes, transformed features, and output results, enabling model inspection from multiple angles”[19].

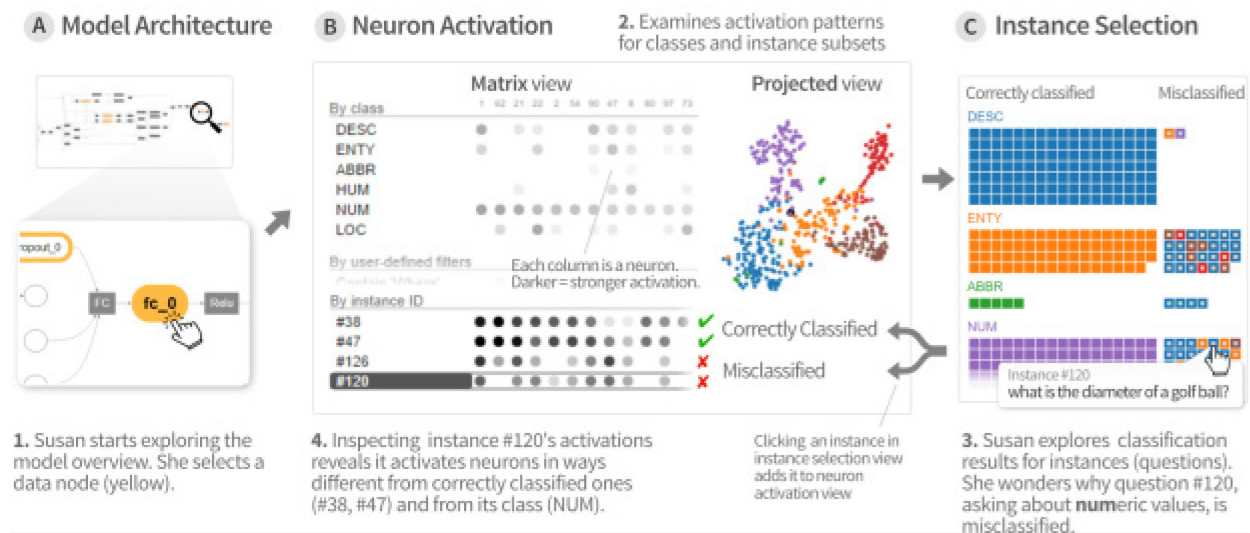


Figure 6-5: ActiVis Integrates Several Coordinated Views to Support Exploration of Complex Deep Neural Network Models, at Both Instance and Subset-Level. Reproduced with permission from the author. The original can be found in Ref. [19].

It can be interesting to explore the internal characteristics on deep neural networks as they are being developed. DeepEyes [20] is a progressive VA system that supports the design of neural networks during training. With DeepEyes, one can detect stable layers, degenerated filters, inputs that are not activating any filter; reason on the size of a layer; and decide whether additional layers are required to better classify data.

Training deep generative models (DGMs) is more complex than other types of deep models and requires more skill, experience, and know-how. DGMTracker [21] is a VA tool for better understanding and diagnosing that training process. A large amount of time series data that represents training dynamics is extracted and a blue-noise polyline sampling scheme selects time series samples, which can both preserve outliers and reduce visual clutter. To further investigate the root cause of a failed training process, DGMTracker uses a credit assignment algorithm that indicates how other neurons contribute to the output of the neuron causing the training failure. Because a DGM usually contains a CNN or a multilayer perceptron (MLP) as its base component, DGMTracker can be directly used to analyze other types of deep models.

GANViz [22] is a VA tool that is dedicated to help experts understand the adversarial process of Generative Adversarial Networks (GANs) in-depth. It evaluates the model performance of two subnetworks of GANs and allows comparative analysis. The same authors propose another specialized VA tool called DQNViz [23] which focusses on analyzing the Deep Q-Network (DQN), which is one type of deep reinforcement learning model, targets to train an intelligent agent that acquires optimal actions while interacting with an environment.

6.3 EXPLAINING DEEP LEARNING RESULTS

Impressive feats were performed by deep neural networks in the last few years. However, these models remain mostly black box systems and it can be difficult to understand the rationale behind the model’s outputs. DL is required to build the trust needed to take the machine’s recommendations into account, especially in the defence intelligence community where decisions need to be auditable and explainable. Senior leadership expects its staff to be in a position to explain its decisions, it is reasonable to expect AI to do the same if it is to support analysts. Similar concerns brought the European Union to implement a “right to explanation” in their regulation so that users can ask for an explanation of algorithm decisions that apply to them [24].

Interpretation of machine learning models is often linked to identifying the features that underlie the model’s predictions and their relative importance. Some of the visualization of the node activations from the previous section can certainly contribute to pointing out what aspect of the input data led to a particular conclusion, i.e., an understanding the inner workings of the network will allow explaining the results.

In addition to these approaches, other researchers trained deep neural networks to provide at least partial explanations as part of the model outputs through an explanation interface. DARPA has initiated the Explainable Artificial Intelligence program (Figure 6-6) for this purpose [25].

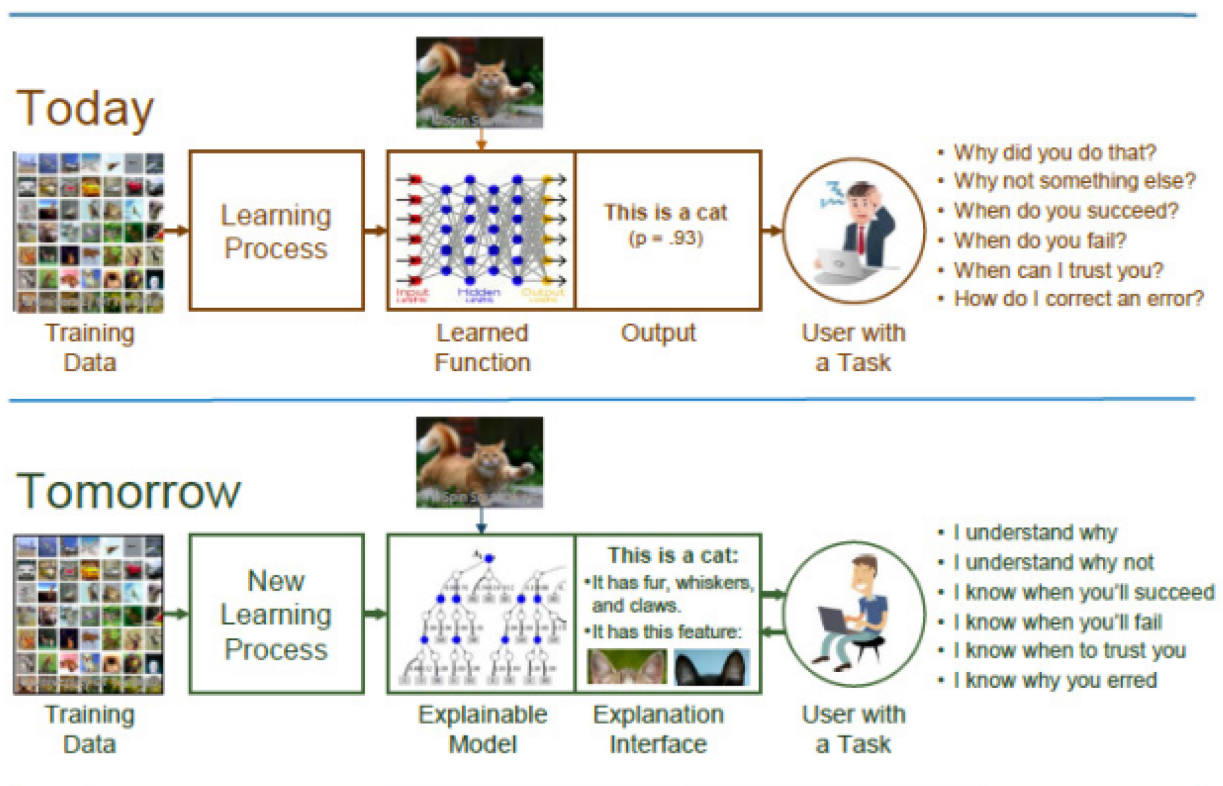


Figure 6-6: Explainable Artificial Intelligence Concept as Presented by DARPA XAI. Reproduced with permission from the author. The original can be found in Ref. [26].

The Local Interpretable Model-agnostic Explanations (LIME) technique [27] defines explanations as models which can be readily presented to the user with visual or textual artifacts, in order to assess trust. With LIME they show that even non-experts are able identify irregularities and perceive when a prediction seems correct but is made for the wrong reasons (see Figure 6-7).

The Layer-wise Relevance Propagation (LRP) framework [28] was proposed for decomposing predictions of modern AI systems in terms of their input variables. Figure 6-8 explains their concept for explaining predictions while Figure 6-9 provides examples of explanations for different types of data and tasks.



Figure 6-7: Left: Husky Classified as Wolf, Right: Explanation of the Model's Prediction in the "Husky vs Wolf" Task. The prediction is correct but was caused by the presence of snow in the background. Reproduced with permission from the author. The original can be found in Ref. [27].

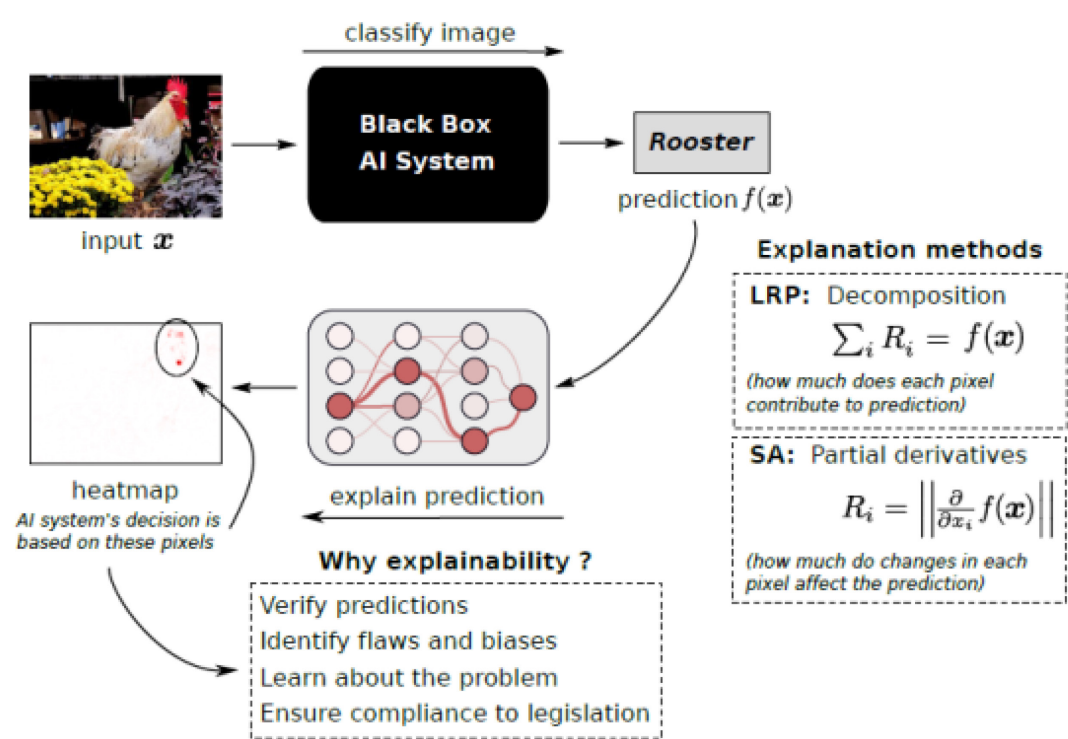
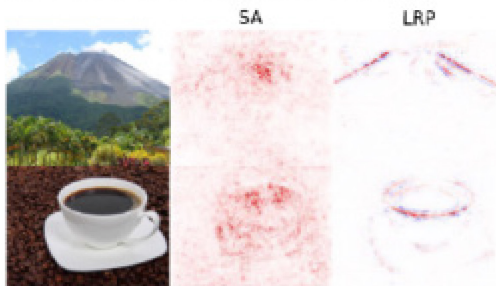


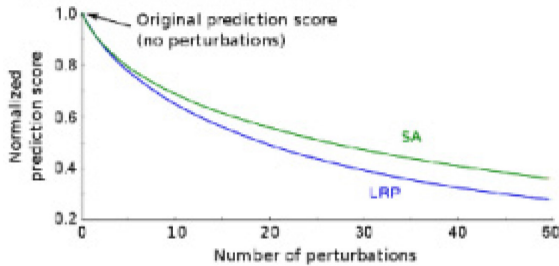
Figure 6-8: The Input Image is Correctly Classified as "Rooster". In order to understand why the system has arrived at this decision, explanation methods such as SA or LRP are applied. The result of this explanation is an image, the heatmap, which visualizes the importance of each pixel for the prediction. In this example, the rooster's red comb and wattle are the basis for the AI system's decision. Reproduced with permission from the author. The original can be found in Ref. [28].

(A) Image classification

Explaining predictions: "Volcano", "Coffe Cup"

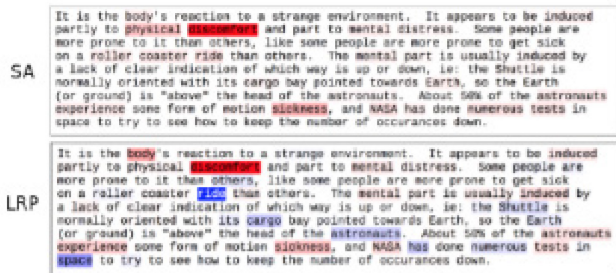


Quantitative comparison of SA and LRP

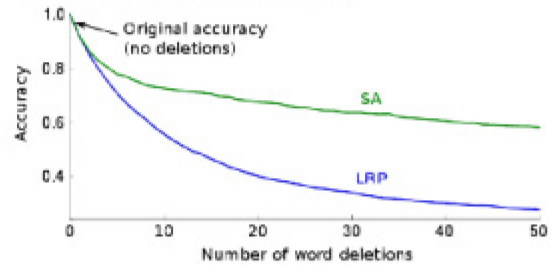


(B) Text document classification

Explaining prediction: "sci.med"



Quantitative comparison of SA and LRP



(C) Human action recognition in videos

Explaining prediction: "sit-up"

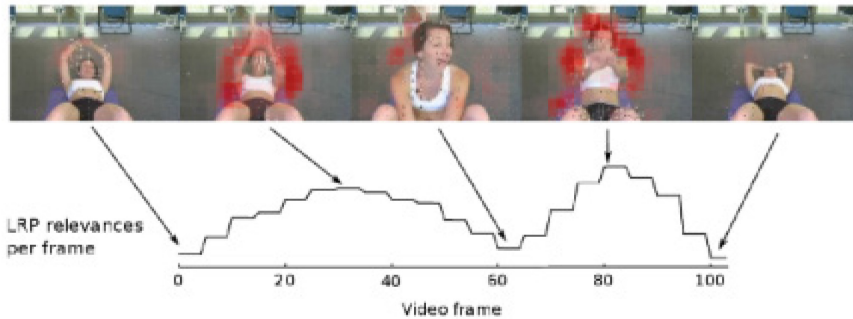


Figure 6-9: Explaining Predictions of AI Systems. (A) shows the application of explainable methods to image classification. SA heatmaps are noisy and difficult to interpret, whereas LRP heatmaps match human intuition. (B) shows the application of explainable methods to text document classification. The heatmaps identify words such as “discomfort,” “body” and “sickness” as the relevant ones for explaining the prediction “sci.med.” In contrast to sensitivity analysis, LRP distinguishes between positive (red) and negative (blue) relevances. (C) shows explanations for a human action recognition classifier based on motion vector features. LRP heatmaps of a video which was classified as “sit-up” show increased relevance on frames in which the person is performing an upwards and downwards movement. Reproduced with permission from the author. The original can be found in Ref. [28].

The generation of visual explanations [29] that clearly states the rationale for a classification has a very high potential and these neural networks could become part of VA systems as well. This approach is advantageous because it does not require the user to have familiarity with the underlying AI system. Figure 6-10 shows the architecture of the model and Figure 6-11 provides some visual explanations generated with this system.

EluciDebug [30] is proposed as the first interactive system designed to help end users build useful mental models of a machine learning system while allowing them to explain back corrections to the system. This is an instantiation of the idea of “Explanatory Debugging” proposed by Kulesza et al. [31], where the user is

identifying and correcting the systems’ faults in a two-way exchange of explanations between the end user and a machine. This prototype is not based on a deep neural network, but the design principles detailed in the paper could certainly be applied to this type of machine learning too.

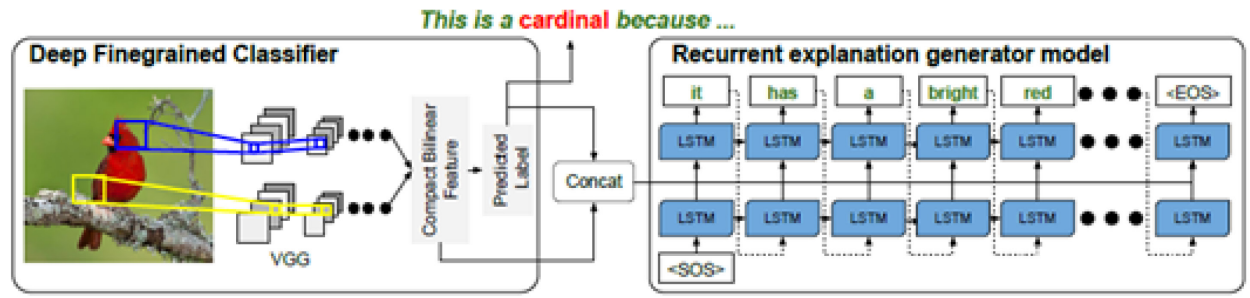


Figure 6-10: Joint Classification and Explanation Model Architecture [29].



Figure 6-11: Visual Explanations Generated by the Model, Containing Image Relevant Sentences with Class Discriminative Attributes. Reproduced with permission from the author. The original can be found in Ref. [29].

6.4 EXPLOITING THE SYNERGY BETWEEN VISUAL ANALYTICS AND DEEP LEARNING

There are many tools that show how VA and DL can be coupled to create powerful applications. Here, the intent is not merely to apply VA to support DL, but rather to leverage both VA and DL to perform complex tasks.

In the study urban environments, Li et al. developed a VA tool [32] that employs DL to help users evaluate the similarities and dissimilarities of urban districts within and across cities. They propose an interactive analysis of the relative importance of a hierarchy of learned visual features across clusterings of different data points.

In the health domain, ECGLens [33] is an interactive system for arrhythmia detection and analysis using large-scale ECG data. It integrates automatic heartbeat classification by a CNN model, an outlier detection algorithm, and interactive visualization techniques. The same approach could be of interest for classification and outlier detection in other types of military-relevant signals.

Another type of data that is relevant to military operations are space-time data, or “tracks.” A visualization methodology for mobility data using artificial neural net techniques is proposed in Ref. [34]. The use a Self-Organizing Map (SOM), to cluster the visiting patterns, a RNN for predicting time series analysis and interactive visualization to enable the user to interpret the results and better understand the behavior patterns of moving objects (see Figure 6-12).

Generative adversarial networks are another kind of deep neural network that can be leveraged in interactive applications. Although not strictly a VA approach per se, the interactive image manipulation tool [35] shown in Figure 6-13 illustrates the potential of integrating generative DL with interactive visualization for information retrieval capabilities.

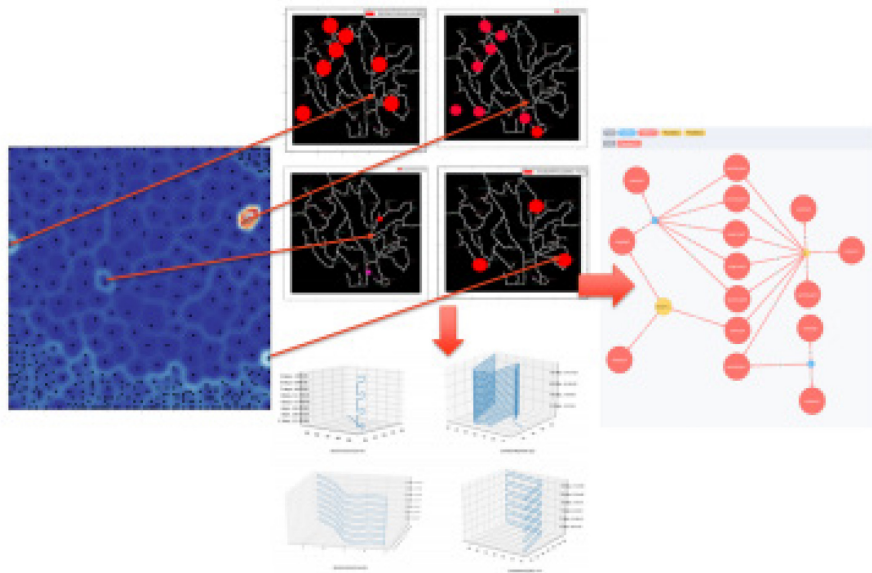


Figure 6-12: Exploring Clustering Result of VAST 2017, There are Four Outstanding Patterns and Such Patterns Visualized by Heat Map, 3D Map and Relationship Map. Reproduced with permission from the author. The original can be found in Ref. [34].

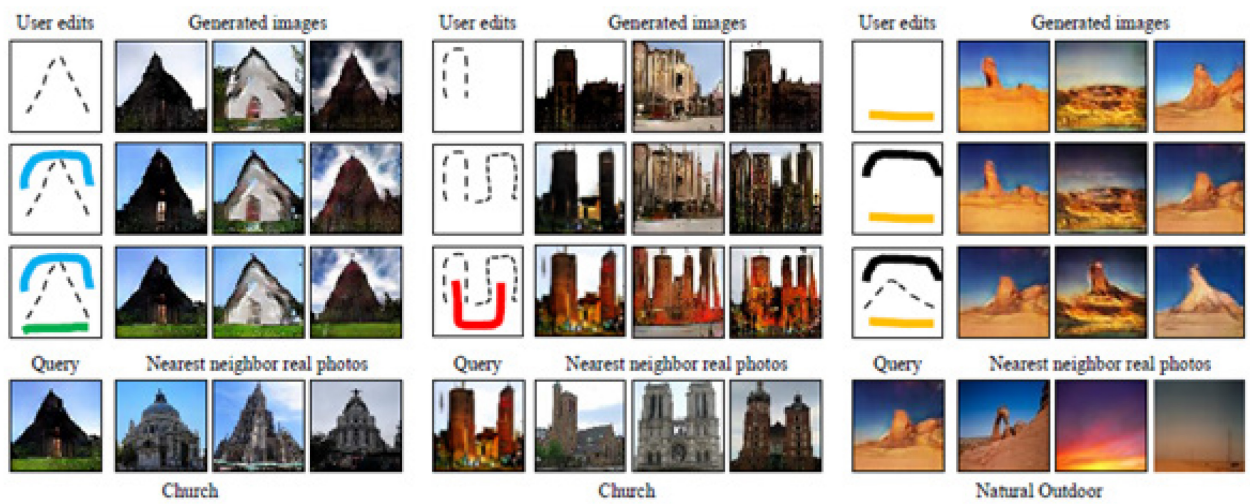


Figure 6-13: Interactive Image Generation. The user uses the brush tools to generate an image from scratch (top row) and then keeps adding more scribbles to refine the result (2nd and 3rd rows). In the last row, we show the most similar real images to the generated images (dashed line for the sketch tool, and color scribble for the color brush). Reproduced with permission from the author. The original can be found in Ref. [35].

6.5 CONCLUSION

In this study, we explored how applications at the intersection of VA and DL techniques can provide a better understanding of deep neural networks and their results, as well as enable powerful applications. Approaches that combine VA to AI will be a key driver of future defence and security systems with advanced processing capabilities required by increasing task complexity, but where one cannot just act on AI predictions in blind faith.

6.6 REFERENCES

- [1] Wang, J. (2019). Interpreting and Diagnosing Deep Learning Models: A Visual Analytics Approach. Electronic Thesis or Dissertation, <https://etd.ohiolink.edu/>.
- [2] Hohman, F., Kahng, M., Pienta, R., and Horng Chau, D. (2018), Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers. arXiv:1801.06889v1.
- [3] Choo, J., and Liu, S. (2018). Visual Analytics for Explainable Deep Learning. IEEE Computer Graphics and Applications, 2018, arXiv:1804.02527.
- [4] Endert, A., Ribarsky, W., Turkey, C., Wong, W., Nabney, I., Díaz Blanco, I., and Rossi, F. (2018). The State of the Art in Integrating Machine Learning into Visual Analytics. arXiv:1802.07954v1.
- [5] Lu, Y., Garcia, R., Hansen, B., Gleicher, M., and Maciejewski, R. (2017). The State-of-the-Art in Predictive Visual Analytics, Computer Graphics Forum, 36: pp. 539-562, doi:10.1111/cgf.13210.
- [6] Tensorflow Playground, <https://playground.tensorflow.org>. Accessed on May 1, 2020.
- [7] Karpathy, A. ConvNetJS, <https://cs.stanford.edu/people/karpathy/convnetjs/>. Accessed on May 1, 2020.
- [8] TensorBoard, <https://github.com/tensorflow/tensorboard>. Accessed on May 1, 2020.
- [9] Wongsuphasawat, K., Smilkov, D., Wexler, J., Wilson, J., Mané, D., Fritz, D., and Wattenberg, M. (2018). Visualizing Dataflow Graphs of Deep Learning Models in TensorFlow. In IEEE Transactions on Visualization and Computer Graphics, 24(1), pp. 1-12.
- [10] Visdom, <https://github.com/facebookresearch/visdom>. Accessed on May 1, 2020.
- [11] DL4J Training UI, <https://deeplearning4j.org/visualization>. Accessed on May 1, 2020.
- [12] Zeiler, M.D., and Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In European Conference on Computer Vision 2014, Springer, pp. 818-833.
- [13] Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., and Lipson, H. (2015). Understanding Neural Networks Through Deep Visualization. Deep Learning Workshop, 31st International Conference on Machine Learning, Lille, France, 2015. arXiv:1506.06579.
- [14] Chung, S., Suh, S., Park, C., Kang, K., Choo, J., and Kwon, B.C. (2016). ReVACNN: Real-Time Visual Analytics for Convolutional Neural Network, KDD'16 Workshop on Interactive Data Exploration and Analytics.
- [15] Chung, S., Park, C., Suh, S., Kang, K., Choo, J., and Kwon, B.C. (2016). ReVACNN: Steering Convolutional Neural Network via Real-Time Visual Analytics. In Future of Interactive Learning Machines Workshop at the 30th Annual Conference on Neural Information Processing Systems (NIPS), 2016.
- [16] Maaten, L.V.D., and Hinton, G. (2008). Visualizing Data Using t-SNE. Journal of Machine Learning Research, 9(Nov), pp. 2579-2605.
- [17] Ming, Y., Cao, S., Zhang, R., Li, Z., Chen, Y., Song, Y., and Qu, H. (2017). Understanding Hidden Memories of Recurrent Neural Networks. IEEE Conference on Visual Analytics Science and Technology (IEEE VAST 2017). arXiv:1710.10777.

- [18] Strobel, H., Gehrmann, S., Pfister, H., and Rush, A.M. (2016). LSTMVis: A Tool for Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks. arXiv:1606.07461.
- [19] Kahng, M., Andrews, P.Y., Karlo, A., and Chau, D.H. (2017). ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models. arXiv:1704.01942.
- [20] Pezzotti, N., Höllt, T., Van Gemert, J., Lelieveldt, B.P.F., Eisemann, E., and Vilanova, A. (2018). DeepEyes: Progressive Visual Analytics for Designing Deep Neural Networks. TVCG, 24(1), pp. 98-108.
- [21] Liu, M., Shi, J., Cao, K., Zhu, J., and Liu, S. (2018). Analyzing the Training Processes of Deep Generative Models. IEEE TVCG, 24(1), pp. 77-87.
- [22] Wang, J., Gou, L., Yang, H., and Shen, H. W. (2018). GANViz: A Visual Analytics Approach to Understand the Adversarial Game. IEEE Transactions on Visualization and Computer Graphics, 24(6), pp. 1905-1917.
- [23] Wang, J., Gou, L., Shen, H. W., and Yang, H. (2018). DQNViz: A Visual Analytics Approach to Understand Deep q-Networks. IEEE Transactions on Visualization and Computer Graphics, 25(1), pp. 288-298.
- [24] Goodman, B., and Flaxman, S. (2016). European Union Regulations on Algorithmic Decision-Making and a Right to Explanation. Presented at 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY. arXiv:1606.08813.
- [25] DARPA, Explainable Artificial Intelligence (XAI), Program Information, <https://www.darpa.mil/program/explainable-artificial-intelligence>.
- [26] Gunning, D. (2016). Explainable Artificial Intelligence (XAI). IJCAI 2016 Workshop on Deep Learning for Artificial Intelligence (DLAI), 2016.
- [27] Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. arXiv:1602.04938.
- [28] Samek, W., Wiegand, T., and Müller, K.-R. (2017). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. arXiv:1708.08296.
- [29] Hendricks, L.A., Akata, Z., Rohrbach, M., Donahue, J., Schiele, B., and Darrell, T. (2016). Generating Visual Explanations. arXiv:1603.08507.
- [30] Kulesza, T., Burnett, M., Wong, W.-K., and Stumpf, S. (2015). Principles of Explanatory Debugging to Personalize Interactive Machine Learning. IUI 2015, March 29 – April 1, 2015, Atlanta, GA, USA.
- [31] Kulesza, T., Stumpf, S., Burnett, M.M., Wong, W.-K., Riche, Y., Moore, T., Oberst, I., Shinsel, A., and McIntosh, K. (2010). Explanatory Debugging: Supporting End-User Debugging of Machine-Learned Programs. In Proceedings of the 2010 IEEE Symposium on Visual Languages and Human-Centric Computing 2010, pp. 41-48.
- [32] Li, L., Tompkin, J., Michalatos, P., and Pfister, H. (2017). Hierarchical Visual Feature Analysis for City Street View Datasets. IEEE VIS 2017 Workshop on Visual Analytics for Deep Learning, October 2017.

- [33] Xu, K., Guo, S., Cao, N., Gotz, D., Xu, A., Qu, H., Yao, Z., and Chen, Y. (2018). ECGLens: Interactive Visual Exploration of Large Scale ECG Data for Arrhythmia Detection. Proc. the ACM CHI Conference on Human Factors in Computing Systems, 2018.
- [34] Chen, Z., Zhou, J., and Wang, X. (2017). Visual Analytics of Movement Pattern Based on Time-Spatial Data: A Neural Net Approach. arXiv:1707.02554.
- [35] Zhu, J., Krähenbühl, P., Shechtman, E., and Efros, A.A. (2016). Generative Visual Manipulation on the Natural Image Manifold. arXiv:1609.03552.



Chapter 7 – CYBER SITUATION AWARENESS

Margaret Varga

University of Oxford
UNITED KINGDOM

Carsten Winkelholz and Susan Träber-Burdin

FKIE
GERMANY

Petter Bivall

Swedish Defence Research Agency
SWEDEN

Kaur Kullman

Cognitive Data OÜ
ESTONIA

7.1 INTRODUCTION

We have become more and more dependent on the ever-expanding Internet with its growing complexities and inter-dependencies. While on the one hand we benefit from its immensely powerful infrastructure, we are vulnerable to cyber-attacks which can happen anytime, anywhere and can cause widespread service degradation and network destruction [1], [2], [3], [4].

In 2008, NATO set up the NATO Cooperative Cyber Defence Centre of Excellence (NATO CCDCOE) in Tallinn [3]. The strategic concept document of NATO 2020 states – responding to the rising danger of cyber-attacks:

NATO must “accelerate efforts to respond to the danger of cyber-attacks by protecting its own communications and command systems, helping Allies to improve their ability to prevent and recover from attacks, and developing an array of cyber defence capabilities aimed at effective detection and deterrence.” [5].

Cyber Situation Awareness (SA) is vital in support of making informed decisions for maintaining a stable, safe and secure environment [6], [7]. The enhancement of the cyber operators’ situation awareness is thus a crucial objective for any user interface design [8], [9], [10], [11], [12].

In addressing parts of the challenges associated with cyber SA the IST-141 group organized an inter-Panel specialists’ meeting, IST-HFM-154, on cyber symbology, in Dayton, Ohio (2016). The aim of the meeting was to gather experts, users, and stakeholders to discuss and progress the development of symbols used for cyber information in a NATO context.

This chapter summarizes the work conducted in the IST-141 group regarding cyber SA and cyber symbology, including studies conducted by the IST-141 group in exploring, developing, and comparing user-centered and system-based approaches to facilitate cyber SA.

7.2 CYBER SITUATION AWARENESS

Endsley’s work on SA provides an established definition of SA, in particular for dynamic environments:

Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future [13].

Endsley considered that there are three stages of SA, namely: 1) Perception; 2) Comprehension; and 3) Projection [13]. This links cognitive psychology with human factors (see Chapter 2: Human Factors Considerations for Visual Analytics) in the sense making process, and this is particularly important in complex situations such as the cyber domain. In cyber operations Endsley's three stages of SA are closely interlinked and are undertaken in round the clock operations. The necessary awareness covers the network infrastructure in both the physical and the virtual domains. The analysts need to be able to detect, recognize, identify, and communicate trends, patterns, violations, and anomalies in an intuitive and timely manner [5].

Indeed, cyber SA is concerned with the human cognitive process and the processing of data. In the complex and dynamic cyber environment, the quality and the speed of human decision making can be greatly enhanced by acute situation awareness. This chapter reports our study on exploring, developing, and comparing different human machine interface design approaches applicable in addressing the analysis and presentation as well as the use of cyber symbology for cyber SA.

7.3 HUMAN MACHINE INTERFACE DESIGN APPROACHES

Different approaches to human machine interface design can be developed and applied to address the different operational and users' needs, for example:

- User-centered approaches; and
- System-based approaches, such as the Ecological Interface Design (EID).

These two approaches provide SA in a different manner. The user-centered approach is concerned with the users' and the tasks' needs, the users' skills, and limitations, as well as their mental models [8], [14]. While, the EID focuses on the system [15], [16], [17], [18], [19], [20], [21] with the aim to show the complex relationships in the system to the user in a readily informative manner. It is a user interface design particularly suitable for real-time dynamic and complex socio-technical systems [18], [21], [22].

Figure 7-1 and Figure 7-2 show examples of user-centered and system-based approaches for cyber security, details of the work can be found in Refs. [23], [24], [25].

An initial evaluation by analysts was carried out for the user-centric and the EID approaches developed in this study. It was found that the user-centric visualization approach provided an effective means of analyzing, detecting, discovering, and identifying patterns, anomalies, violations, and threats; as well as correlating events, Figure 7-1. The resulting intuitive visualizations are suitable for the provision of detailed information on the performance of network components, such as IPs, ports, protocols, packages, CPU load, disk, and memory usages, etc. The EID visualization, on the other hand, portrayed the logical network topology, the functionalities of the network, depicting the relationships and dependencies between servers, firewalls, etc. It provided a visualization that guided the users to understand the functioning of the network. Once users became familiar with the patterns of the 'normal situation,' they could readily detect any changes from the normal patterns, Figure 7-2. Therefore, in the EID concept, analysts easily saw the operational aspects of the network, i.e., the big picture. The two approaches complement each other in providing awareness and information of different aspects of the network situation [26].

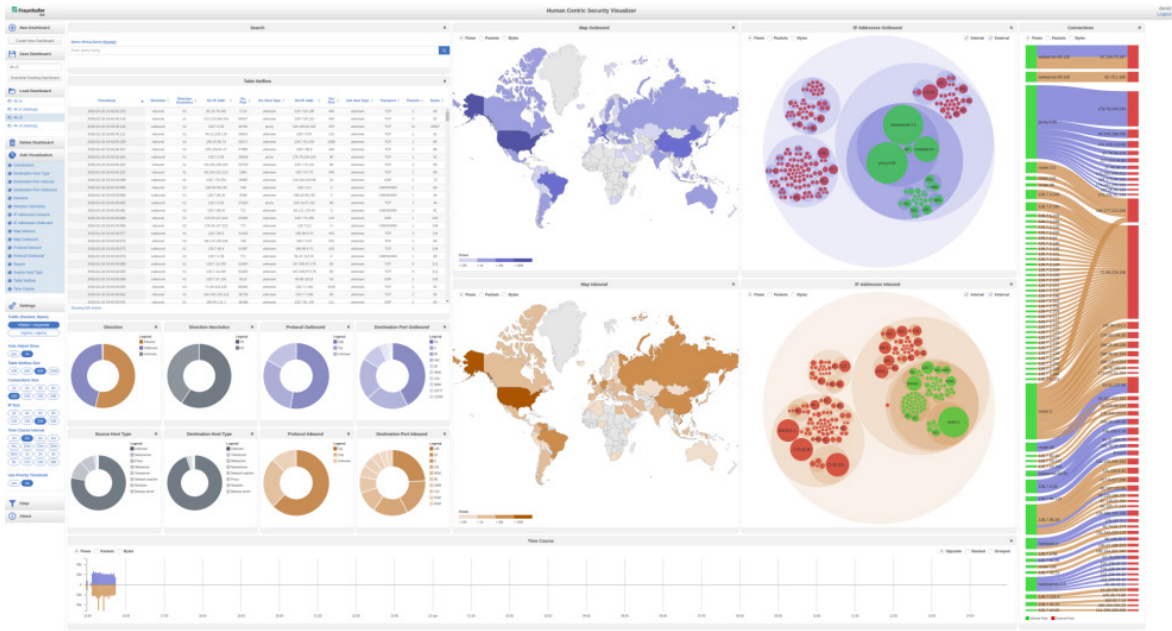


Figure 7-1: User-Centric Approach.

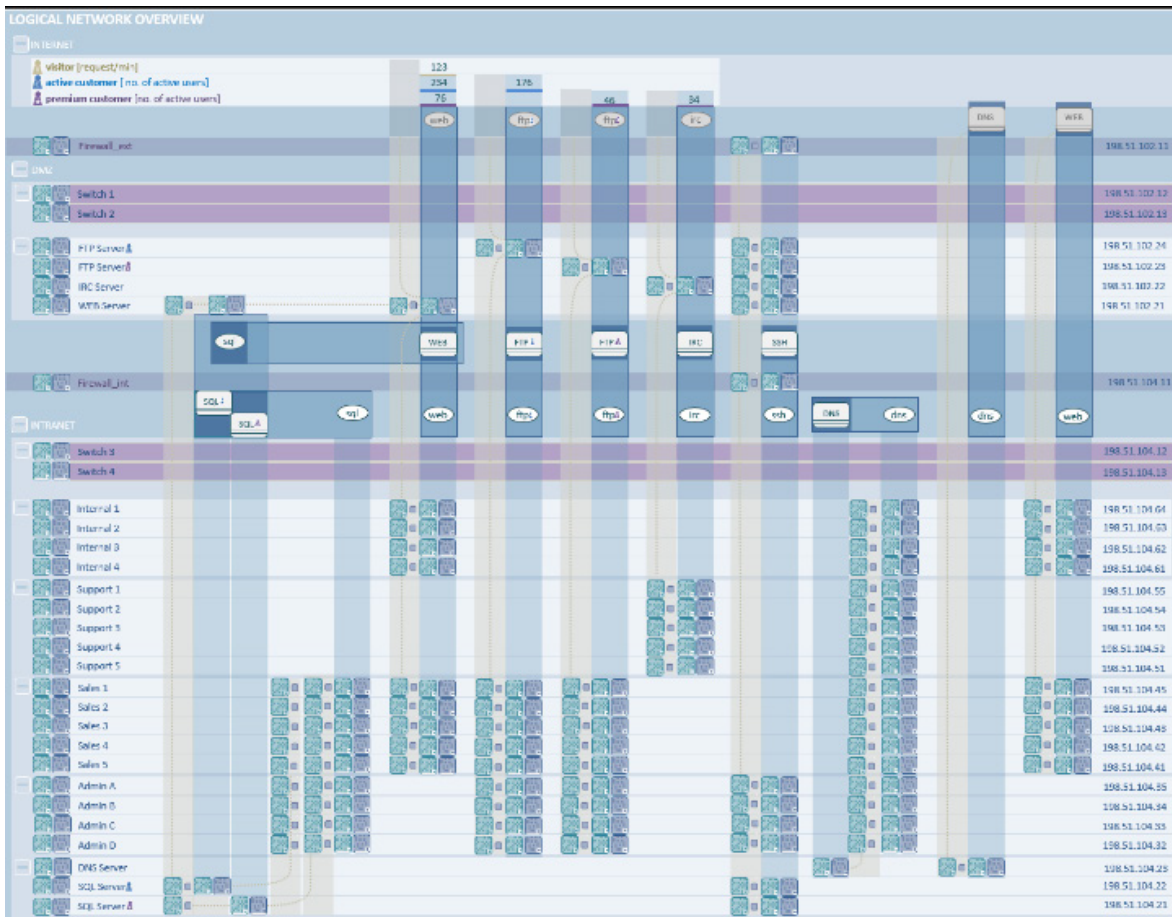


Figure 7-2: System-Based – Ecological Interface Design.

7.4 SYMBOLOGY

Some symbols were created in the above system-based approaches. What is a symbol? A symbol is a representation of an object, idea, emotion, relationship or thought, etc. It can be composed of a combination of words, gestures, sounds, ideas, or visual images. However, users need to learn what they mean and how to associate the symbols with their meaning before they can use them. For example, CO₂ (Carbon dioxide) is an acidic colorless gas with a molecular structure composed of one carbon atom and two oxygen atoms. Some suggest that symbols can be made more intuitive and easier to remember by using pictorial or iconic representations.

In general, symbols that represent physical objects are easy to construct and use; but there is not a natural physical representation for abstract representations of tasks, ideas, actions, thoughts, applications, etc. Symbols can thus be complex, but the key advantage of using a symbol is that it can be given very specific meaning and has great potential as an information carrier. A set of symbols can be created for specific purposes, domains, or applications, and in the case of the cyber domain, to provide an intuitive means of accurately conveying, for example, a networks' status and to make it easier for different users with different backgrounds and levels of expertise to understand the situation of the network.

7.4.1 Symbology, Geo-Spatial and Cyber-Spatial Thinking

Traditionally military operations have always been carried out on the ground, in the air or space, on the water surface or below. In all cases, it has been possible to keep track of the units, activities and events using symbols on a map, that is, the geo-spatial aspect has always been a prominent part of the information flow. The symbol standards used in the United States and by NATO [13], [28], [29], have been developed with these conditions in mind, and with adaptations suitable for military fallback approaches, such as the requirement that all symbols should be possible to draw by hand directly on a map. Cyber, on the other hand, does not follow the geo-spatial boundaries and carries a very different set of challenges when visualized.

Military operations in the geo-spatial world have fairly clear boundaries; an airplane cannot travel indefinitely far in a matter of seconds, a certain number of troops can be expected to be spread out over a limited space, a missile only reaches targets within its range, etc. Cyber is connected to hardware executing instructions and mediating data, but cyber-spatial operations have little to no restrictions with respect to space or time as the structure is more based on activities and abilities of the computational hardware [30], processor microcode [31], peripherals firmware [32], various layers of software [33], [34], and last, but not least, the cyber operators. A server can be located anywhere in the world, activities can move from one country to another in seconds, and automated systems can survey an adversary's network continuously for months. When working with cyber, these differences require a move from geo-spatial thinking to cyber-spatial thinking. How this transition should be achieved and how the visual representations and cyber symbology are to be designed to support the cyber-spatial thinking is a topic for both ongoing and future research. Figure 7-3 aims to illustrate the move from geo-spatial to cyber-spatial using two of the many different types of representations available.

Additionally, the manner in which users interpret symbols, and thus become aware of the situation, its effect and impact, differs a lot between geo-spatial and cyber-spatial representations (symbology). To complicate things further, there are also cases when geo-spatial and cyber-spatial merge, raising a need to relate the cyber situation to missions and operations in the physical space as well as in the logical space [28]. The following is our list of challenges adapted from Refs. [23], [35]:

- a) How do icons compare with symbols in conveying the required information?
- b) When to use symbols and when to use icons and in what operations/activities/domain(s)/layers ?
- c) How to decide what is the intuitive way to depict multiple cyber elements and their associated situations?

- d) How to decide what is the intuitive way to depict multiple cyber situations?
- e) How much information and detail are necessary for and from different user groups and different operational needs?
- f) How to show the temporal elements of a situation?
- g) Should cyber symbols be superimposed on geo-spatial maps and / or should other aspects such as the network architecture be included?
- h) Scalability.
- i) How can user (system) performance be evaluated?
- j) User identification and requirement capturing.

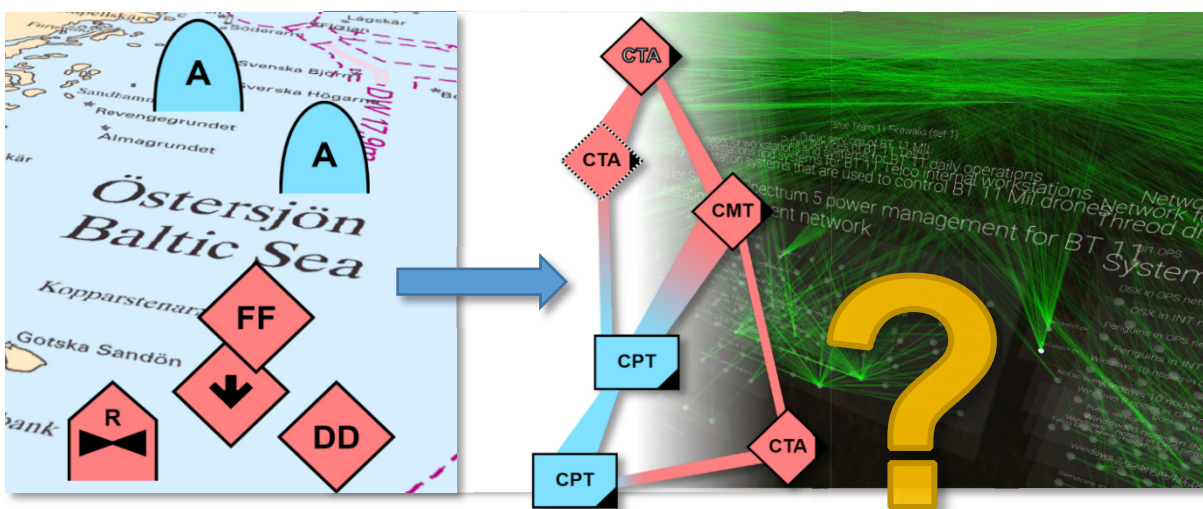


Figure 7-3: Moving from Geo-Spatial to Cyber-Spatial Visual Representations is a Requirement for Successful Cyber SA. How the cyber-spatial representations should be designed is still a topic for research. Left: Traditional geo-spatial symbols; right: alternatives for cyber-spatial representations, including an example from Ref. [36].

7.4.2 MIL-STD-2525D and NATO APP-6

Military Standard (MIL-STD)-2525D is a major document defining the rules and requirements for the development and display of joint military symbology in the US within the Department of Defense (DoD) and non-DoD entities across all services and functions. It is mainly concerned with geographic-centric representation of the physical layer and covers a very wide range of military symbology applications and requirements and includes civilian unit / organization symbols [37]. In June 2014, an initial set of cyber symbols were added to MIL-STD- 2525D through the addition of Appendix L [27].

NATO Standardization Agreement (STANAG) Allied Procedural Publication (APP) 6 is the joint NATO Standard on Military Symbology for maps. The symbols provide a standard set of common operational symbols (Command, Communications and Control Measures symbols) which have been designed to enhance NATO’s interoperability in joint and combined (different forces Air/Land/Sea/etc., units, organizations, and nationalities) military operations. It can be used in electronic / automated display systems and manual applications. The symbology consists of graphical symbols for visual representation of physical objects, activities, or events. The symbols are applicable in multi-chrome and monochrome and can be hand drawn. APP-6 also recognizes the need for cyberspace symbology [29].

APP-6 and MIL-STD-2525 aim to develop a comprehensive joint military symbology that is common to both organizations, which covers physical (units, equipment), non-physical (planning) or predicted locations with temporarily assigned characteristics or validity [27], [29], [38]. A ‘building block’ approach is used to represent their symbology: an icon-based symbol is used to depict units, equipment, installations, activities, and meteorological occurrences, etc. which are located within and around a virtual bounding octagon concept. The octagon can be composed of 1) Frames; 2) Icons; 3) Modifiers and 4) Amplifiers, as well as color, graphics, and alphanumeric representations. Such symbols can be assimilated by users much more readily and effectively than a text (language) based communication, which can be prone to misunderstanding, misinterpretation and imprecision, ambiguity as well as slowness of transmission [37].

The main advantage of these two standards is that they are familiar to the military, thus the incorporation of cyber symbols can be readily accepted; the disadvantages are the adaptation / extension of physical symbols into the non-physical aspects of the cyber domain is complex and may result in misinterpretation.

7.4.3 Other Cyber Symbology Approaches

McCroskey and Mock [39] developed a MIL-STD-2525 compliant symbol set that addresses the lack of effective communication between the physical and cyber domains, having highlighted the risk that decisions on strategic and mission planning will be made without awareness of the network situation, and thus could endanger warfighters. Their MIL-STD-2525 compliant symbol set is concerned with the display of cyber information in the Logical and Persona layers of cyberspace. The advantage of their proposed MIL-STD 2525D compliant cyberspace operational graphics approach is that it is more likely to be accepted by the existing military users. It provides a means to convey cyber information that is relevant to commanders who are unfamiliar with the technical aspects of the cyberspace that affects their decision making. The disadvantage is that it follows a ‘pre-cyber’ symbology approach which could bound the way symbols could be developed, and therefore, also, the need to show how the physical and logical networks relate to each other.

Examples of standardized geo-spatial symbols and non-standardized cyber-spatial symbols are presented in Figure 7-4.







Geo-spatial	Description	Cyber-spatial	Description
	Friendly airborne fixed wing.		Shellcode action, a remote control action.
	Hostile airborne rotary wing.		Scanner action, actions aimed at mapping out a network or gaining information about computers.
	Friendly surface vessel.		Exploit action, attacks aimed at acquiring privileges on computers.

Figure 7-4: Examples of MIL-STD-2525D Geo-Spatial Symbols (Left) and Cyber-Spatial Symbols Created Independently in the Absence of a Standard (Right). Cyber symbols designed by Jennifer Bedhammar and Oliver Johansson, to be published in a master’s thesis at Linköping University, used under permission.

Varga et al. [23] used spider diagrams as a symbol to represent the dynamics of all the parameters on a map of an internal network (an intranet) for each internal host. This provides an intuitive means for the user to visualize the problems readily as they arise, and thus make informed decision to mitigate the problems [23].

In addition to showing the variation (deviation from norm) of parameters for each host, spider diagrams also show characteristic shapes for different states, and these may be recognized easily by an experienced operator. Such visualizations provide a first step in providing diagnostics information. The advantage of this approach is that it provides an effective means of communicating the dynamic status of the logical network which is applicable in both the military and civil domains. The disadvantage is that there is a need to map the different spider diagrams with diagnostic information. However, this approach is non-compliant with any existing Standards, therefore acceptance into the military operation is not straightforward.

Fugate and Gutzwiller [40] also identified that the characteristics of cyberspace operations differ from those in the physical domain, and it is therefore problematic to regard cyberspace representation as a subset of the physical domain operational picture. They consider that re-using the physical space for cyberspace (MIL-STD-2525D) makes it difficult for users familiar with the physical domain, and associated representation, to differentiate the two, i.e., when cyberspace information is treated as physical information it could lead to misinterpretation and hence mis-informed decision making. While their approach is inspired by MIL-STD-2525 they did not restrict their approach to using physical domain symbology to depict cyber effects and actions for cyber threats. They consider a cyber-attack incident reflects the attacker's approach and motivation and designed three symbols to represent three different entities in a cyber incident, namely, devices, users, and software. They represent a device by a square, while a circle represents a user and a hexagon represents software. Vulnerability in a device is indicated by a broken outline. Each entity is also associated with a trust element, which can be unknown, trusted, untrusted, threat or insider. They also address the scalability issue. The advantage of their approach is it can represent a cyber incident based on the three entities. The disadvantage is that, once again, their approach is not compliant with any existing Standards, and therefore adoption into the military domain is not straightforward.

It can be seen, from the above, that there are some possible approaches towards some of the identified cyber symbology challenges, e.g., d) and f). Many of the challenges remain completely open and unaddressed.

7.5 CONCLUSIONS

This chapter discussed the user-centered and system-based approaches to providing different types, levels and perspectives of cyber SA. Cyber symbology was also discussed.

An initial evaluation found that the user-centric approach to SA provides an effective means of analyzing, detecting, discovering, and identifying patterns, anomalies, violations, and threats, as well as correlating events. The visualizations are suitable for the provision of detailed information on the performance of network components.

The EID approach, on the other hand, provides an effective visualization to guide users in their understanding of how networks should function compared to how these networks are actually functioning; thus, analysts can easily see the operational aspects of the network, i.e., the big picture.

The two approaches complement each other in providing awareness and information on different aspects of the network situation.

Cyber symbology has the potential to enable visualization of cyber situation, though there are not yet clear methodologies or solutions as to how it can best be achieved. This chapter lays out cyber symbology challenges

and questions that must be answered by future research; thus, providing a foundation and context for future research programs to develop military cyber symbology.

7.6 REFERENCES

- [1] Ablon, L., and Bogart, A. (2017). Zero Days, Thousands of Nights; The Life and Times of Zero-Day Vulnerabilities and Their Exploits. Santa Monica: RAND Corporation. Retrieved from https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1751/RAND_RR1751.pdf
- [2] Kott, A., Wang, C. and Erbacher, R.F. (Eds.) (Jan 2015). Cyber Defense and Situational Awareness. Springer.
- [3] Geers, K. (2011). Strategic Cyber Security, CCD COE Publication, ISBN 978-9949-9040-7-5.
- [4] Peng, T., Leckie, C., and Ramamohanarao, K. (2007). Survey of Network-Based Defense Mechanisms Countering the Dos and DDos Problems. ACM Computing Survey, 39(1), 3.
- [5] NATO 2020: Assured Security; Dynamic Engagement, 17th May 2010. https://www.nato.int/cps/en/natolive/official_texts_63654.htm.
- [6] D’Amico, A., and Whitley, K. (2008). The Real Work of Computer Network Defense Analysts: The Analysis Roles and Processes that Transform Network Data Into Security Situation Awareness. In Proceedings of the Workshop on Visualization for Computer Security (VizSec 2007), Springer, Berlin, pp. 19-37.
- [7] Franke, U., and Brynielsson, J. (2014). Cyber Situational Awareness – A Systematic Review of the Literature. Computer and Security, 26, pp. 18-31, Elsevier.
- [8] Endsley M.R., and Jones, D.G. (2004). Designing for Situation Awareness: An Approach to User Centered Design, Second Edition, CRC Press, ISBN 9781420063554.
- [9] Grégoire, M. and Beaudoin, L. (2004). Visualisation for Network Situational Awareness in Computer Network Defence. NATO IST-043 Visualisation and the Common Operational Picture, Toronto, Sept 2004.
- [10] Lahmadi, A., and Beck, F. (2015). Powering Monitoring Analytics with ELK Stack. 9th International Conference on Autonomous Infrastructure, Management and Security, June 2015, Ghent, Belgium.
- [11] Lavigne, V., and Gouin, D. (2014). Visual Analytics for Cyber Security and Intelligence. The Journal of Defence Modeling and Simulation: Applications, Methodology, Technology, April 2014. <http://dms.sagepub.com/content/11/2/175>.
- [12] Varga, M.J., Winkelholz, C., Träber-Burdin, S., and Bivall, P. Cyber Situation Awareness, International conference of Cyber Defence, 13th – 14th April, 2018, Sofia, Bulgaria. (invited).
- [13] Endsley, M.R. Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors, 37(1), pp. 32-64, March 1995.
- [14] McKenna, S., Staheli, D., and Meyer, M. (2015). Unlocking User-Centered Design Methods for Building Cyber Security Visualizations. IEEE Symposium on Visualization for Cyber Security (VizSec), 2015.

- [15] Bennett, K.B. (2014). VEILS: An Ecological Interface for Computer Network Defense. Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting, 2014.
- [16] Burns, C.M., Kuo, J., and Ng, S. (2003). Ecological Interface Design: A New Approach for Visualizing Network Management. *Computer Networks* 43, pp. 369-388, Elsevier.
- [17] Burns, C.M. (2000). Putting It All Together: Improving Display Integration in Ecological Displays. *Human Factors* 42, pp. 224-241.
- [18] Rasmussen, J., and Vicente, K.J. (1989). Coping with Human Errors Through System Design: Implications for Ecological Interface Design. *International Journal of Man-Machine Studies*, 31, pp. 517-534.
- [19] Rasmussen, J. (1985). The Role of Hierarchical Knowledge Representation in Decision Making and System Management. *IEEE Transactions on Systems, Man and Cybernetics*, 15, pp. 234-243.
- [20] Rasmussen, J. (1983). Skills, Rules, Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man and Cybernetics*, 13, pp. 257-266.
- [21] Rasmussen, J. (1974). The Human Data Processor as a System Component. *Bits and Pieces of a Model*. Risø National Library, Risø-M No. 1722.
- [22] Vicente, K., and Rasmussen, J. (1992). Ecological Interface Design: Theoretical Foundations, *IEEE Transactions on Systems, Man and Cybernetics* 22, pp.1-18.
- [23] Varga, M.J., Winkelholz, C., and Träber-Burdin, S. (2019). An Exploration of Cyber Symbology. *IEEE VizSec*, 23rd Oct 2019, Vancouver, Canada.
- [24] Varga, M.J., Winkelholz, C., and Träber-Burdin, S. (2018). An Exploration of User Centered and System Based Approaches to Cyber Situation Awareness. 15th IEEE Symposium on Visualization for Cyber Security (VizSec), 22nd Oct 2018, Berlin, Germany.
- [25] Varga, M.J., Winkelholz, C., and Träber-Burdin, S. (2018). Exploration of User Centered and System Based Approaches to Cyber Situation Awareness. NATO HFM-288 Research Workshop on Integrated Approach to Cyber Defence: Human in the Loop, 16th – 18th Apr 2018, Sofia, Bulgaria.
- [26] Varga, M.J., Winkelholz, C., and Träber-Burdin, S. (2017). Exploration of User Centered and System Based Approaches to Cyber Situation Awareness. 14th IEEE Symposium on Visualization for Cyber Security (VizSec), 2 Oct 2017, Phoenix, USA.
- [27] Department of Defense (2014). Interface Standard Joint Military Symbology, MIL-STD-2525D, 10 June 2014. M.R., Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32-64, March 1995.
- [28] Joint Publication 3-12, Cyberspace Operations, 8 June 2018. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_12.pdf Accessed 26 May 2019
- [29] NATO. (2011). Allied Procedure Publication (APP) – 6(c) NATO Joint Military Symbology, May 2011. [https://web.archive.org/web/20150921231042/http://armawiki.zumorc.de/files/NATO/APP-6\(C\).pdf](https://web.archive.org/web/20150921231042/http://armawiki.zumorc.de/files/NATO/APP-6(C).pdf)
- [30] Intel. (n.d.). Intel Product Security Center Advisories. <https://www.intel.com/content/www/us/en/security-center> Accessed 02 May 2020.

- [31] Bitdefender (02 Oct 2020). LVI-LFB Side-Channel Attack. <https://www.bitdefender.com/business/cyber-threats/lvi-lfb-attack.html> Accessed 28 Apr 2020.
- [32] Cojocar, L., Kim, J., Patel, M., Tsai, L., Saroiu, S., Wolman, A., and Mutlu, O. (2020). Are We Susceptible to Rowhammer? An End-to-End Methodology for Cloud Providers. 41st IEEE Symposium on Security and Privacy (S&P). Internet, 2020. Retrieved from <https://www.microsoft.com/en-us/research/publication/are-we-susceptible-to-rowhammer-an-end-to-end-methodology-for-cloud-providers/>
- [33] Carnegie Mellon University. (n.d.). Beware of Compiler Optimizations. Retrieved from <https://wiki.sei.cmu.edu/confluence/display/c/MS06-C.+Beware+of+compiler+optimizations>. Accessed on April 28, 2020.
- [34] MITRE Corporation. (n.d.). ATT&CK Matrix for Enterprise. <https://attack.mitre.org/> Accessed 28 Apr 2020.
- [35] Varga, M.J., Winkelholz, C., Träber-Burdin, S., Liggett, K., Werner, K., Bivall, P., and Lavigne, V.A (2016). A Consideration of the Application of Icons and Symbols in Cyber Situation Awareness. NATO IST-HFM-154 Cyber Symbology Specialists' Meeting, 28th – 30th Nov 2016, Dayton, USA.
- [36] Kullman, K., Ryan, M., and Trossbach, L. (2019). VR/MR Supporting the Future of Defensive Cyber Operations, NATO Computer Aided Analysis, Exercise, Experimentation Forum, 2019.
- [37] Andrews, M., and Loveridge, S. (2016). Joint Symbology Standard Management in the Military Domain. NATO IST-HFM-154 Specialists' Meeting on Cyber Symbology, 28th – 30th Nov 2016.
- [38] McGrane, B., Bohling, J., and Eple, M. (2016). Development, Distribution and Management of a Common Cyber Symbology for Joint Military Planning and Operations. NATO IST-HFM-154, Cyber Symbology Specialists' Meeting, 28th – 30th November 2016, Ohio, USA.
- [39] McCroskey, E.D. and Mock, C.A. (2017). Operational Graphics for Cyberspace. Joint Force Quarterly 85, 2nd Quarter, 2017.
- [40] Fugate, S.F., and Gutzwiller, R.S. (2016). Rethinking Cyberspace Symbology. NATO IST-HFM-154, Cyber Symbology Specialists' Meeting, USA, November 2016.

Chapter 8 – IMPROVISED EXPLOSIVE DEVICE INCIDENTS ANALYSIS WITH STORYTELLING EXPLORATORY VISUAL ANALYTICS

Valérie Lavigne
Defence Research and Development
CANADA

Shivas Jayaram
BrainCradle
UNITED STATES

Marius Panga
Microsoft
UNITED STATES

8.1 INTRODUCTION

The collection and analysis of datasets about incidents can help analysts derive activity level assessment, perform trend analysis, and gain greater insight about the problem they are monitoring. In order to better understand the dynamics of Improvised Explosive Device (IED) attacks and support work on Counter-IED (C-IED), we decided to apply an exploratory visual analytics approach with storytelling elements. Analyzing datasets about incidents can help derive activity level assessment, perform trend analysis, and gain better situation awareness.

Exploratory visual analytics is meant to allow a user to explore a dataset and discover interesting patterns and insights. Some visual analytics tools afford complex interactions that can be intimidating to a new user. Storytelling techniques can help a new user get started with a new interactive visualization application in order to understand what a dataset contains and how to leverage the exploratory visual analytics' tool capabilities to perform new analyses on the dataset.

We found that the use of an exploratory visual analytic tool allowed the discovery of interesting patterns in the Ukraine IED events dataset produced by the NATO Counter-IED Center of Excellence (NATO C-IED COE), and we conveyed those insights to users by applying storytelling techniques as part of the interactive visualization tool. Once the users have started exploring the provided insights, they are invited to continue their exploration allowing them to make additional discoveries from the data.

8.2 DATASETS DESCRIPTIONS

8.2.1 NATO Ukraine IED Incidents Data

This is a NATO Unclassified IED events spreadsheet from the NATO C-IED COE. It contains 665 events, with 15 in 2001-2013, about 230 in 2014 and the rest in 2015. It contains the following column headers: Date, Type, KIA (Killed in Action), WIA (Wounded in Action), City, Region, Country, Details, and Remarks. There are missing values in the dataset.

8.2.2 Ukraine Census Data

We also included various statistics about Ukraine to see if we can find patterns between the IED events and these statistics. The data was cleansed and transformed in order to have the same common attributes as the primary data set (the NATO IED dataset). We applied the following transformations in order to facilitate the data merge with the other datasets: removed irrelevant data from the data set, converted all fields to the correct type, and translated the region names to the ones used by the primary data set.

8.3 EXPLORATORY VISUAL ANALYTICS STORYTELLING TOOL

8.3.1 Design Goal and Approach

This project's goal was to allow the exploration of over 600 IED incidents in Ukraine, mostly over the past 2 years, for the purpose of highlighting and better understanding the temporal, geographical and political patterns in that data. We also considered additional census data and election statistics to uncover potential regional patterns, as well as publicly available data about the existing conflict in Eastern Ukraine. Our intent was to employ interactive visualization to generate better insights about the Ukraine IED situation. We documented how we gathered and transformed the data, our interactive visualization design process and inspirations, and user feedback evaluations in our process book [1].

In order to drill down into different aspects of the dataset, we decided to split it in an overview introducing the purpose of the tool (why) and leading to three specialized views (perspectives) aiming at answering the where, what and how questions about IED incidents.

We repeated the timeline banner and the storytelling boxes at the top of each view and reused the same casualty icons across the visualizations as a way to keep consistency and help user orientation in the tool. Each perspective offers multiple coordinated views that can be filtered temporally using the interactive timeline.

8.3.2 Dataset Overview

The dataset overview approach aligns with the Visual Information-Seeking Mantra of "Overview first, zoom and filter, then details-on-demand" [2]. It introduces the purpose of the visual analytics tool and provides links to the other perspectives (Figure 8-1).

8.3.3 Geospatial View

The geospatial view shown in Figure 8-2 provides insights into the geo-temporal aspects of IED incidents. The user can drill down into each region to see how IED incidents happen over time down to city level. The map display shows how individual incidents are spread over each area, and the incident dots can be colored according to their type or effect on casualties. We also included additional statistical data about Ukraine, allowing regions to be colored by population density or Russian/Ukrainian ethnicity in addition to the incident frequency data.

8.3.4 Incident Type View

In Figure 8-3, the incident types and properties views provide a Sankey diagram that focuses on the incident type and casualties. Brushing over incident types with the mouse displays examples of incident of that type to the right of the diagram.

8.3.5 Text Analysis View

The text analysis in Figure 8-4 view exploits Natural Language Processing (NLP) to analyze the unstructured text descriptions of IED incidents. In the graph, events sharing similar words are linked together and the most prominent words in each group are displayed.

Ukraine
Improvised Explosive Device Incidents

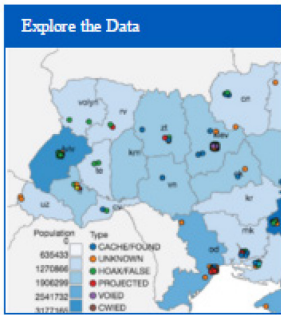
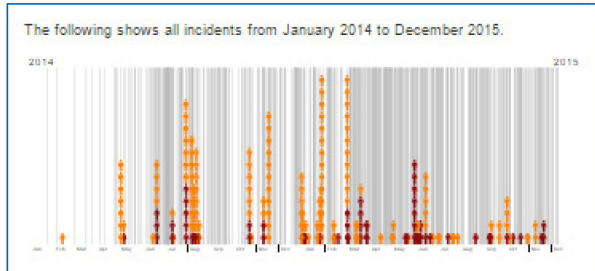
Why Where What How Process

Since the collapse of the Kremlin-supported government in February 2014, Ukraine has been affected by political and economic turmoil.

What followed the turmoil was the emergence of an energized public who would do anything to get their freedom.

We will analyse the various incidents that occurred between January 2014 to December 2015. Our exploration of over 600 Improvised Explosive Devices (IED) incidents in Ukraine over the past 2 years for the purpose of highlighting and better understanding the temporal, geographical and political patterns in the data. The following shows the summary of the incidents that happened.

65 Killed
152 Wounded
650 Incidents



Summary

Background to the conflict
In February 2014, the Kremlin-supported government of Ukraine collapsed. The demise of the regime was brought about by bitter protests over a decision by the government to reject closer relations with the European Union. What followed the turmoil was the emergence of a pro-Western, pro-reform government and an energized public generally anxious to lessen Moscow's influence and committed to addressing the need for serious reforms.

Ukraine's problems are not solely political and economical. Russia responded to the change over government in 2014 by seizing Ukraine's Crimea region and annexing it.

Improvised Explosive Devices (IED)

An IED can be almost anything with any type of material and initiator. It is a "homemade" device that is designed to cause death or injury by using explosives alone or in combination with toxic chemicals, biological toxins, or radiological material. IEDs can be produced in varying sizes, functioning methods, containers, and delivery methods. IEDs can utilize commercial or military explosives, homemade explosives, or military ordnance and ordnance components. They are unique in nature because the IED builder has had to improvise with the materials at hand. Designed to defeat a specific target or type of target, they generally become more difficult to detect and protect against as they become more sophisticated.

IEDs can be hidden anywhere: on animals, planted in roads or strapped to a person. They can be detonated via cell phones or trip wires, among other methods. They can be deployed everywhere: in a combat environment or in the middle of a busy city. The adaptability of IEDs to almost any situation makes them difficult to detect.

IEDs, or Improvised Explosives Devices, are one of the main causes of casualties among troops and exact a heavy toll on local populations.

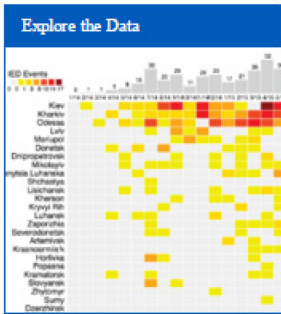
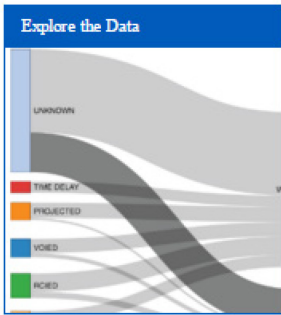


Figure 8-1: C-IED Analysis Tool Introduction View.

IMPROVISED EXPLOSIVE DEVICE INCIDENTS ANALYSIS WITH STORYTELLING EXPLORATORY VISUAL ANALYTICS

Ukraine Improvised Explosive Device Incidents

Why Where What How Process

Where did the Incidents Occur?

Ukraine regions are not equally affected by IED incidents. Incidents were concentrated in three cities: Kiev, Kharkiv and Odessa. There were over 80 incidents in each. However, although there was a much higher number of incidents reported for these cities than for any other city (almost 100), less than 10 persons were killed in those 360 incidents.

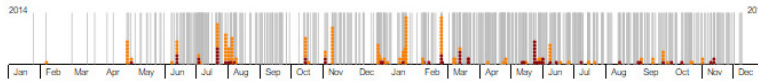
Show Me Kiev

Most of the remaining IED incidents are spread in the Donetsk and the Luhansk regions, representing 136 and 92 incidents respectively. These two areas are also the most active regions for IED incidents at the end of 2015.

Show Me Donetsk

In the reported incidents, many different types of IEDs were mentioned. Some types are more dangerous than others.

Continue to IED Types Exploration ▶



Region color: Population Density
Circle color: IED Type

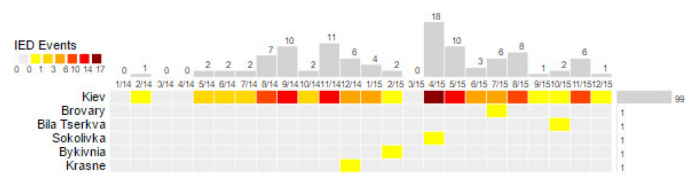
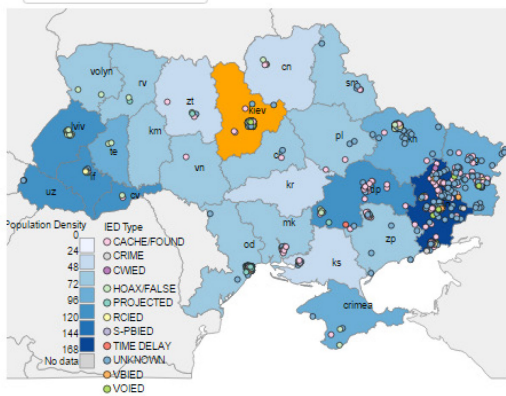


Figure 8-2: C-IED Analysis Tool Geospatial View.

Ukraine
Improvised Explosive Device Incidents

Why Where What How Process

What type of explosives were used?

An IED is a "homemade" device that is designed to cause death or injury by using explosives alone or in combination with toxic chemicals, biological toxins, or radiological material. IEDs can be produced in varying sizes, functioning methods, containers, and delivery methods. They can also be classified in various predefined types in order to aid analysis. Unfortunately, it is often the case that there is not enough information to assign an IED to a specific type.

While 88% of incidents resulted in no casualties, partly because they were hoaxes and false alarms and partly because they were found on time, there is still a significant number of injuries and fatalities. The most dangerous types of IEDs have been identified to be the Remote Controlled devices (RCIEDs), leaving behind 6 dead and 18 wounded, and the Vehicle-borne explosive devices (VBIEDs), leaving behind 12 dead and 12 wounded.

Find out more about how these incidents occurred by going to the next page.

[Continue to Incident Analysis ▶](#)

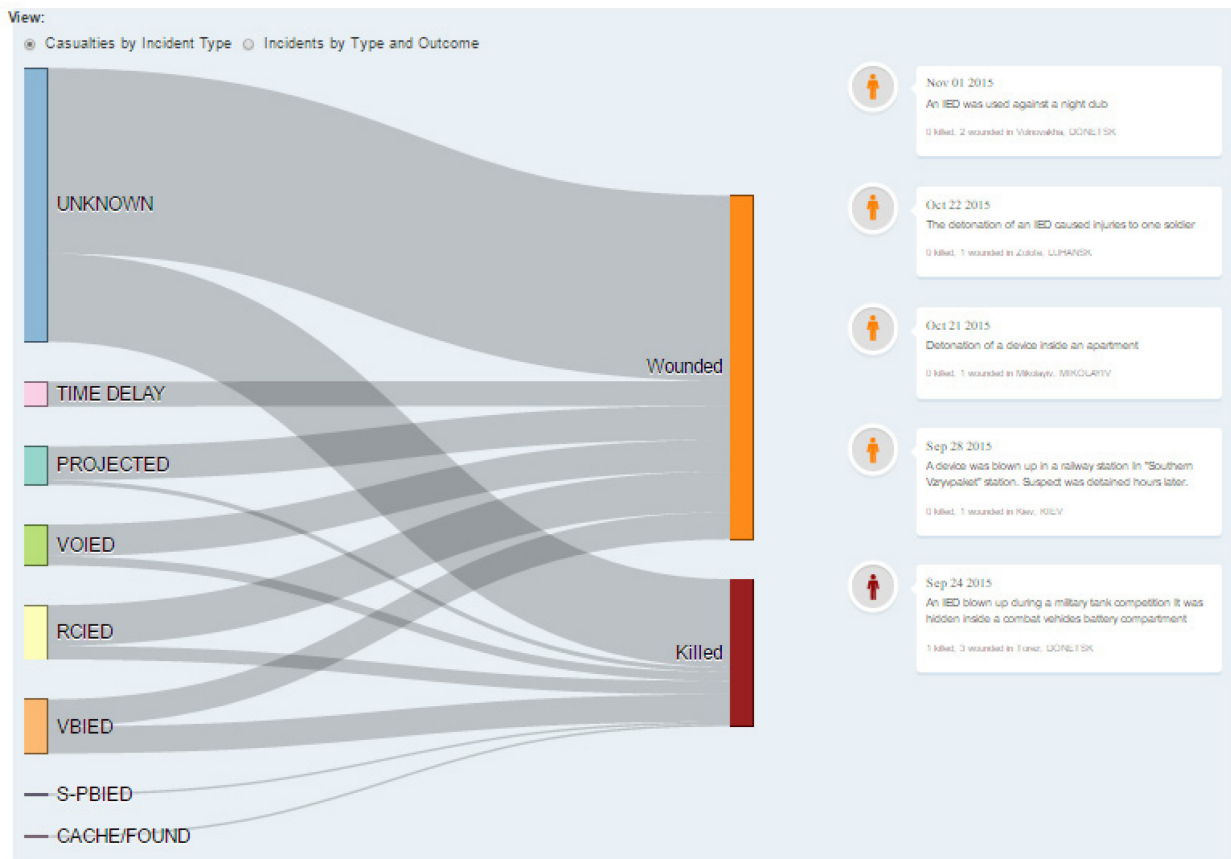
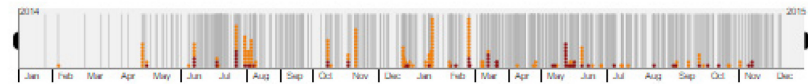


Figure 8-3: C-IED Analysis Tool Incident View.

Ukraine Improvised Explosive Device Incidents	Why	Where	What	How	Process
---	-----	-------	------	-----	---------

How these Incidents Occurred? - Interpreting the Free Text Reports

In this exploration we want to analyze the free text summary written for each incident. For the process of finding report similarity, we first used the TF-IDF and Cosine Similarity algorithms to find similarity between all the incident reports. Then for our analysis we consider two reports to be similar only if their threshold is 50% or higher. To gather important words across report clusters, we first find the three most important words in each of the reports. This was found using TF-IDF (Term Frequency, Inverse Document Frequency) scores. Then for each cluster of nodes we gather all the important words from each node and take the top three scored word. The font size of the words is determined by the frequency of occurrence in the cluster.

By clustering the incident reports in this method we can see that most similar incidents that were reported in metro stations, administration buildings were hoax or false reports which led to no casualties.

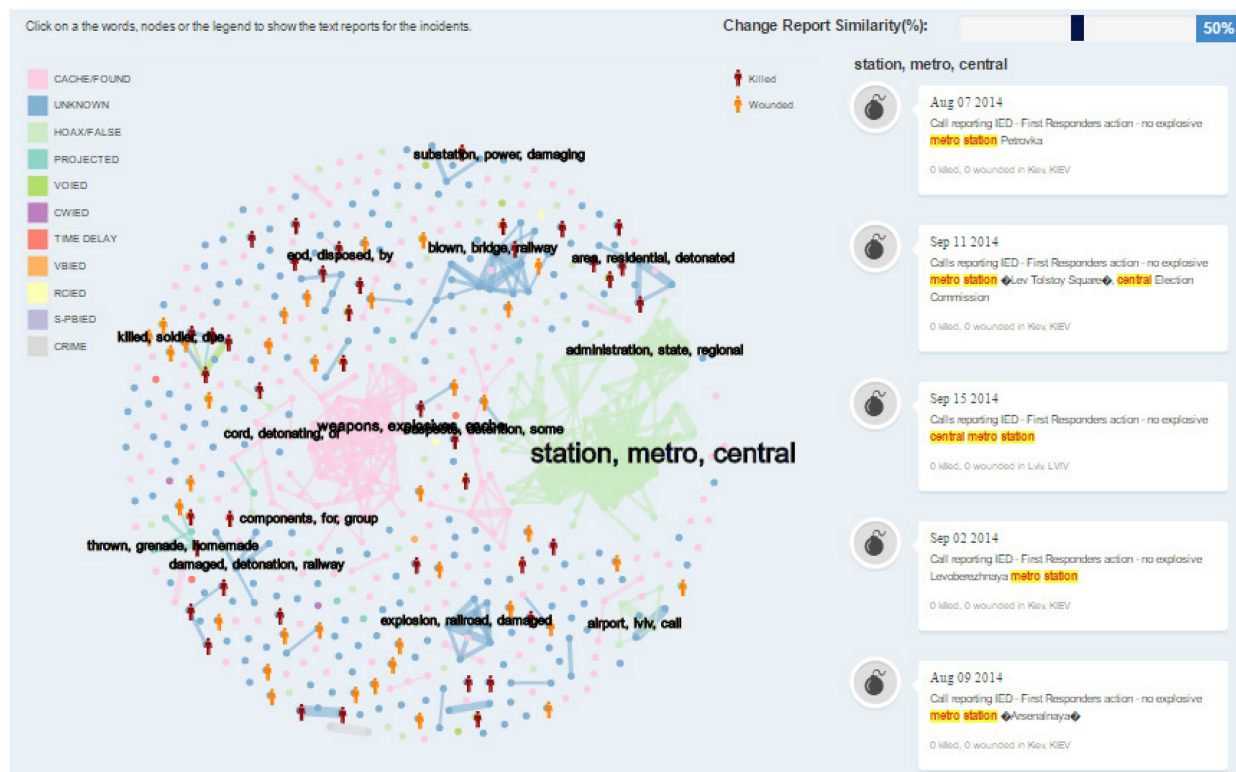


Figure 8-4: C-IED Analysis Tool Text Analysis View.

8.4 STORYTELLING TECHNIQUES

Segel and Heer [3] performed a design space analysis of narrative visualizations. According to this classification, our C-IED analysis tool uses the following three techniques to balance between an author-driven and a reader-driven perspective on the dataset:

- Drill-Down Story;
- Martini Glass Structure; and
- Interactive Slideshow.

8.4.1 Drill-Down Story

The Drill-Down Story visualization structure presents a general theme and then allows the user to choose among particular instances of that theme to reveal additional details and backstories. [3]

In our tool, this is what happens in the overview page (Figure 8-1), which offers a description of the IED context and the incident dataset surrounded by small frames that link to more specialized views to explore this dataset.

8.4.2 Martini Glass Structure

The Martini Glass visualization structure begins with an author-driven approach, initially using questions, observations, or written articles to introduce the visualization. [3]

The Martini Glass Structure is very efficient for showing the user how to use the different features of the interactive visualization tool. The top part of Figure 8-2, Figure 8-3, and Figure 8-4 contains boxes that offer a text providing insights linked to this view of the data and when the user clicks the corresponding “Show Me,” the tool automatically renders the appropriate values and settings to show the insight described with the data.

8.4.3 Interactive Slideshow

The Interactive Slideshow structure follows a typical slideshow format but incorporates interaction mid-narrative within the confines of each slide. This structure allows the user to further explore particular points of the presentation before moving ahead to the next stage of the story. [3]

This pattern appears at the top part of Figure 8-2, Figure 8-3, and Figure 8-4, where the blue boxes discuss particular aspects of the data that stand out in the current view, before offering to continue to the next view. This technique also facilitates navigation between the different views of the tool, along with the top menu.

8.5 INSIGHTS ABOUT IMPROVISED EXPLOSIVE DEVICES INCIDENTS

The dataset about the Ukraine IED situation was produced by the NATO C-IED COE. The classic way to convey information from IED datasets employed by the C-IED COE is through static visualizations and tables that can be incorporated into MS PowerPoint slides, as can be seen from Figure 8-5 and Figure 8-6. Those graphics do not allow an in depth analysis of the dataset and it can be argued that it is not their purpose. Still, it is interesting to explore what additional insights can be extracted from a storytelling visual analytics tool.

Figure 8-5 makes it obvious which regions have the highest number of IED incidents and provides the total number of incidents for the country. We can see that the Kiev (Kyiv) region has a lot of incidents. This is also confirmed in the geospatial view from Figure 8-7. However, if we choose to color the regions according to the lethality of IED incidents (see Figure 8-8), we quickly realize that incidents that happened in Luhansk and Donetsk were more dangerous than the incidents in Kiev and Odesa. The incident dots also show how the incidents are spread in the regions, showing that incidents in Kiev and Odesa are grouped mostly in each region’s capital city, whereas incidents in Luhansk and Donetsk are spread throughout the area.

In Figure 8-6, we can see over time how many incidents happened, and how many persons were either wounded or killed. A pie chart of incident types and a table of events by target type are provided. The interactive timeline banner at the top of each view (shown in Figure 8-9) also provides the information about incidents over time but using a visual representation. The Sankey diagram shows the relative prevalence for each type of incident but adds the information about which ones caused higher casualties. Furthermore, if we focus on casualties, Figure 8-11 shows the effects of each type of IED. Note that Figure 8-9 shows data from

2013, which was not available to us, while Figure 8-10 and Figure 8-11 are based on 2014/2015 data. Using the interactive timeline filter, we can select a sliding window of a few months and discover that hoaxes incidents increase over time. Also, the relative portion of incidents that killed people are increasingly of unknown type.

Finally, the details of IED incidents are not exploited at all in Figure 8-5 and Figure 8-6. Natural language processing allowed us to explore the incident text descriptions to find out that most hoaxes and false IEDs involve the words “station, metro, central.”

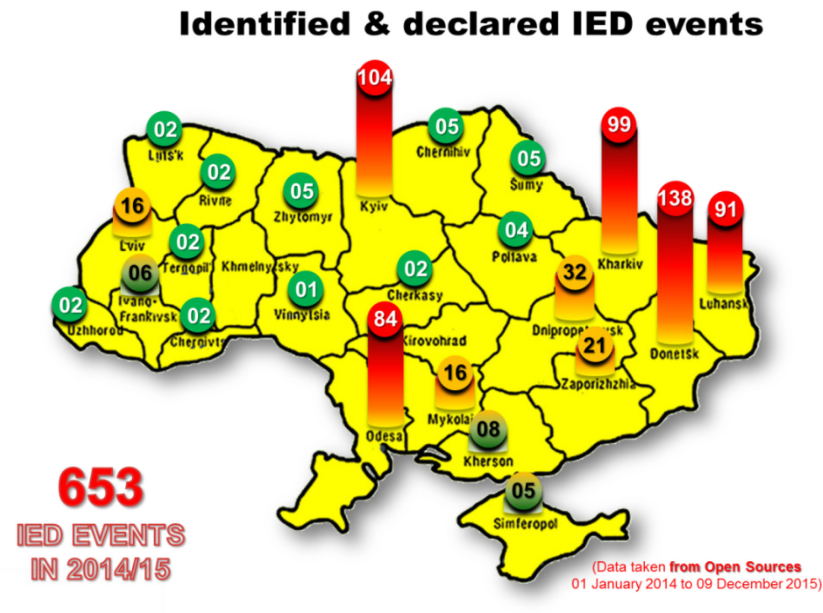


Figure 8-5: Geospatial Summary Slide Featuring a Map of Ukraine IED Incidents in 2014 – 2015.

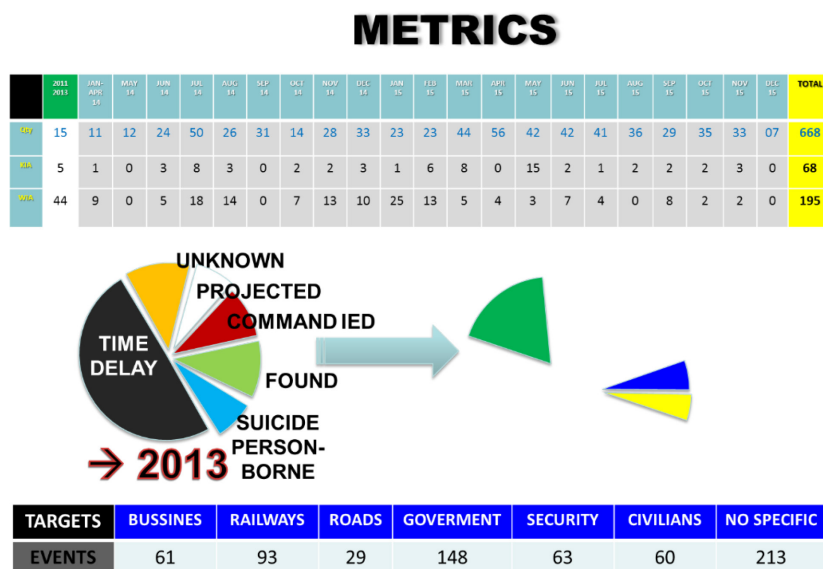


Figure 8-6: Categorical and Temporal Summary Slides Featuring Statistical Data About Ukraine IED Incidents in 2014 – 2015.

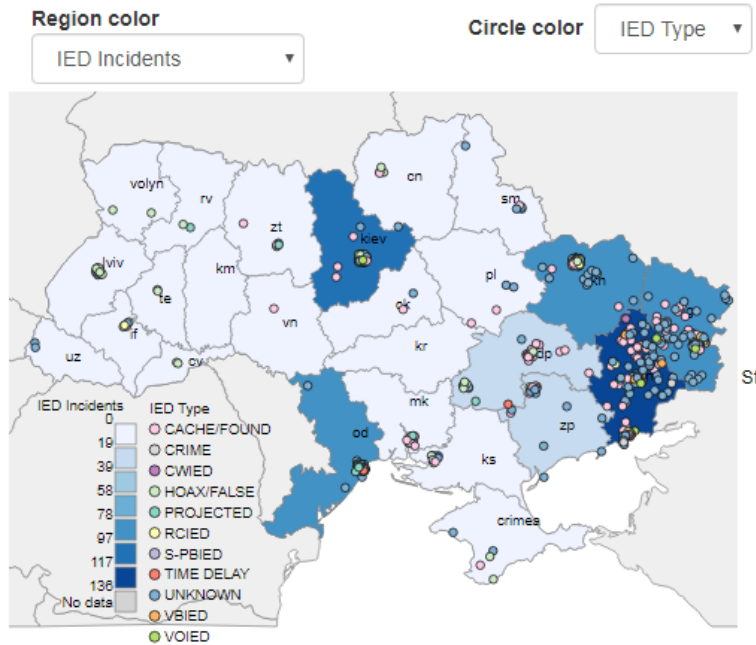


Figure 8-7: Regions Colored According to Level of IED Incidents.

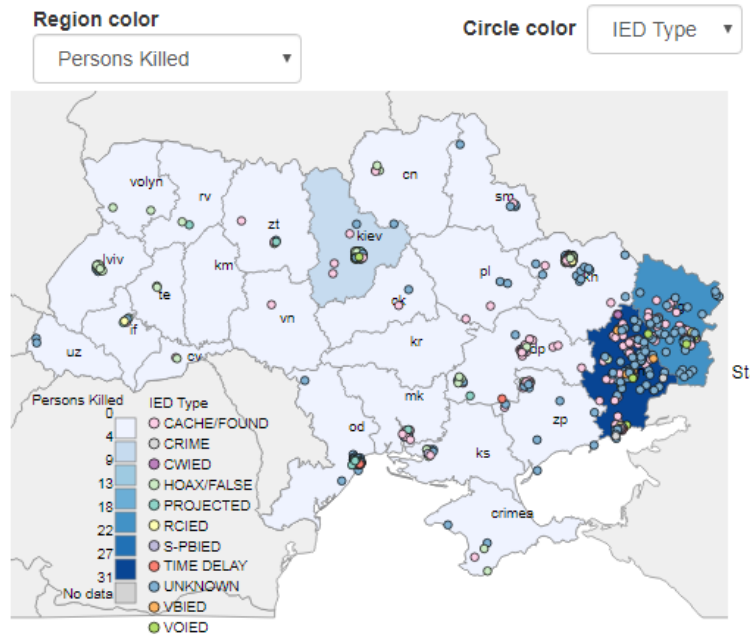


Figure 8-8: Regions Colored According to Lethality of IED Incidents.

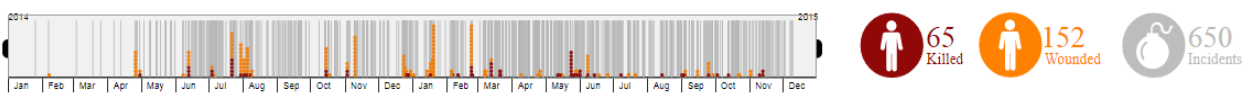


Figure 8-9: Interactive Timeline Banner and Filter at the Top of Each View.

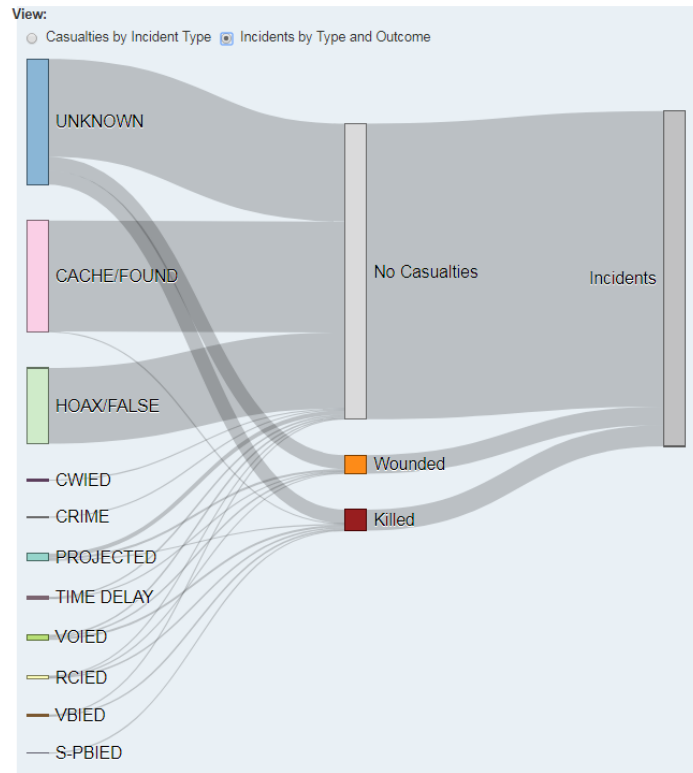


Figure 8-10: Sankey Diagram Showing Incidents by Type and Outcome.

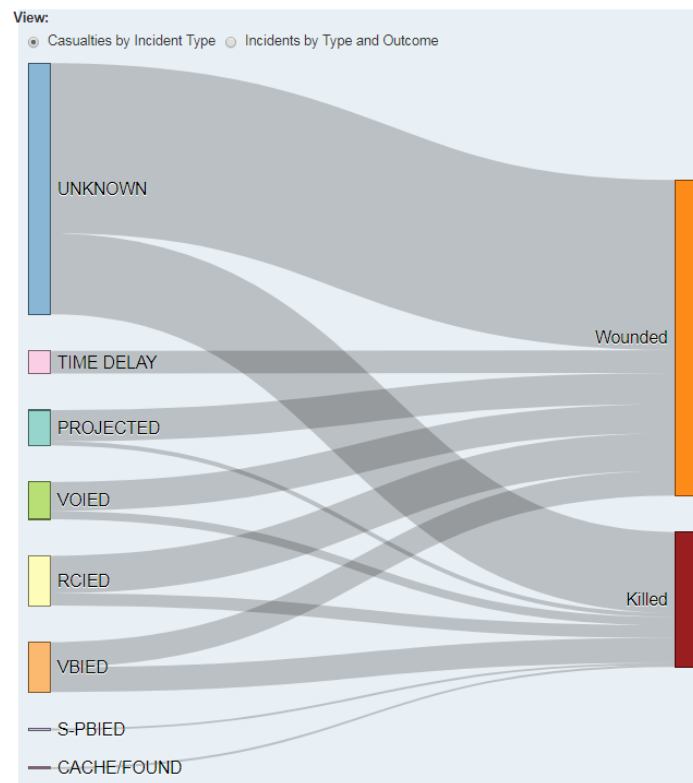


Figure 8-11: Sankey Diagram Showing Casualties by Incident Type.

8.6 CONCLUSION

In this chapter, we described the design approach and storytelling elements behind the exploratory visual analytics tool we developed for analyzing a dataset about IED incidents in Ukraine. We demonstrated that a visual analytics approach can highlight insights that cannot be discovered from summary tables and the basic visualizations provided in presentation slides. In the future, we would recommend improvements to facilitate audit trail considerations, such as a visualization exporting function to save screenshots or interaction videos more easily. The tool is web based and accessible at <http://nato-project.github.io/v2/index.html>. The code for the tool is also freely available. This project was also presented as a positive example of interactive visualization in the “Data Visualization” class at Vytautas Magnus University (Kaunas, Lithuania, 2018 – 2019).

8.7 REFERENCES

- [1] Lavigne, V., Panga, M. and Shivas, J. (2016). Process Book (CS171 Week 13). Project Title: Ukraine Improvised Explosive Devices. http://nato-project.github.io/v2/Process%20_Book.pdf Accessed 9 Sep 2019.
- [2] Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proceedings 1996 IEEE Symposium on Visual Languages, pp. 336-343.
- [3] Segel, E., and Heer, J. (2010). Narrative Visualization: Telling Stories with Data. IEEE Transactions on Visualization and Computer Graphics, 16, pp. 1139-1148.



Chapter 9 – HFM-259 DATA EXPLORATION

Carsten Winkelholz and Nikhil Acharya

Fraunhofer FKIE

GERMANY

9.1 INTRODUCTION

This chapter reports the collaboration between IST-141 RTG “Exploratory Visual Analytics” and HFM-259 RTG “Human Systems Integration Approach to Cyber Security”. They made available their dataset and developed ontology to enable the IST-141 RTG to explore the application of visual analytics to the HFM-259 dataset. This chapter discusses the work conducted in exploring the dataset and ontology.

9.2 DATASET

The HFM-259 RTG “Human Systems Integration Approach to Cyber Security” has developed an ontology to code the 230 scientific papers they identified on the topic of Human Factors in Cyber Security. The ontology combines different concepts from the human factors domain with the cyber security domain [1]. The elements from the ontology have been used to tag papers with the values of the ontology elements. Some of the possible attributes with example values are shown below to provide an overview of the ontology:

- **threat.source**
 - insider
 - outsider
- **threat.motivation**
 - financial gain
 - sabotage
 - ...
- **threat.vectors**
 - malware
 - phishing
 - ...
- **vector.complexity**
 - single
 - multiple
 - ...
- **affected.users**
 - end-users
 - it services
 - ...
- **influencing.techniques**
 - authority
 - scarcity
 - social proof
 - ...
- **information.assurance**
 - availability
 - integrity
 - confidentiality
- **mitigation.mechanism**
 - cyber hygiene training
 - cyber security training
 - ...

- **mitigation.effectiveness**
 - yes
 - no
 - somewhat
- **mitigation.target**
 - end-users
 - it services
 - ...
- **performance.shaping.factors**
 - workload
 - hmi complexity
 - ...
- **situational.context**
 - office
 - home
 - ...
- **human.error.taxonomy**
 - slips
 - lapses
 - ...
- **system.outcome**
 - security
 - usable
 - ...
- **source.methodology**
 - anecdotal
 - case study
 - ...
- **consequences.for.instruments.of.power**
 - military actions
 - economic actions
 - ...

A paper can be tagged with multiple values from each element: for example, the paper can address both end-users and it-staff or can discuss multiple influencing techniques. Possible values for each attribute can be “na” (not applicable) as well as many.

9.3 VISUALIZATION FRAMEWORK

The IvID Framework (Information visualization – Interactive Dashboards) from Fraunhofer FKIE was used to explore the data. It is a generic system for visualizing/exploring document and graph databases. It is similar to other known visual analytics systems like Tableau, Splunk, Kibana but it focuses on using graph databases. In this sense it combines graph visualizations with conventional visualizations. For visual analytics applications, dashboards are interactive and allow combining selections performed in the visualizations into filter requests on the data to produce new views.

Graph databases provide flexible and for most use cases intuitive, transparent data models. In addition, graph databases support operations optimized for graphs, such as determining the shortest paths between nodes in the model, the aggregation of values along paths, and filtering according to attributes on the path. Furthermore, in our experience the linkage of nodes in the model also defines contexts for filtering and can be used for implicit adapting filters in the sense of expectations of the users.

Fraunhofer FKIE uses the Framework to investigate issues of human computer interaction in designing a generic framework for visual analytics dashboards on basis of graph databases.

9.4 DATA PREPROCESSING – BAYESIAN NETWORK

One major point of interest in exploring the data was to examine the co-occurrence of tags which might give indications on how different elements of the ontology interacts and helps to build a model on human factors in cyber security.

To analyze statistically the distributions of values in data sets, Bayesian networks can be used. Bayesian networks represent the factorization of a joint probability distribution with multiple factors into a product of conditional probabilities as a graph, where links indicate statistical dependencies of attributes. Therefore, we applied a Bayesian network structure learning algorithm from the bnlearn package [2] to the data. One challenge was that papers have been tagged with multiple values for most concepts. The solution we choose was to split each concept into its possible values. So that the Bayesian network model corresponding to the joint probability distribution consists of a huge number of nodes. The number of possible nodes in the model results from the sum of the number of values for each concept. We will discuss the implications of this on the quality of the model in the conclusions.

However, in this way a node of the model is binary, meaning if a value of a concept is present in the tags of a paper or not. For structure learning we used a hill climbing algorithm [3].

9.5 ANALYTICS DASHBOARDS

9.5.1 Facet Exploration

To get a quick overview about the tags used from the ontology we assembled a dashboard with tag-clouds for each element in the ontology and arranged them aside of a table listing the 230 papers. The table can be used to get detailed information on papers. We choose for each category a specific color which was also used to indicate the category of a node in the Bayesian network. We tested two layouts for the tag-clouds. One visualized with the circle pack layout (Figure 9-1) from d3.js [4] and one with a horizontal bar layout (Figure 9-2).

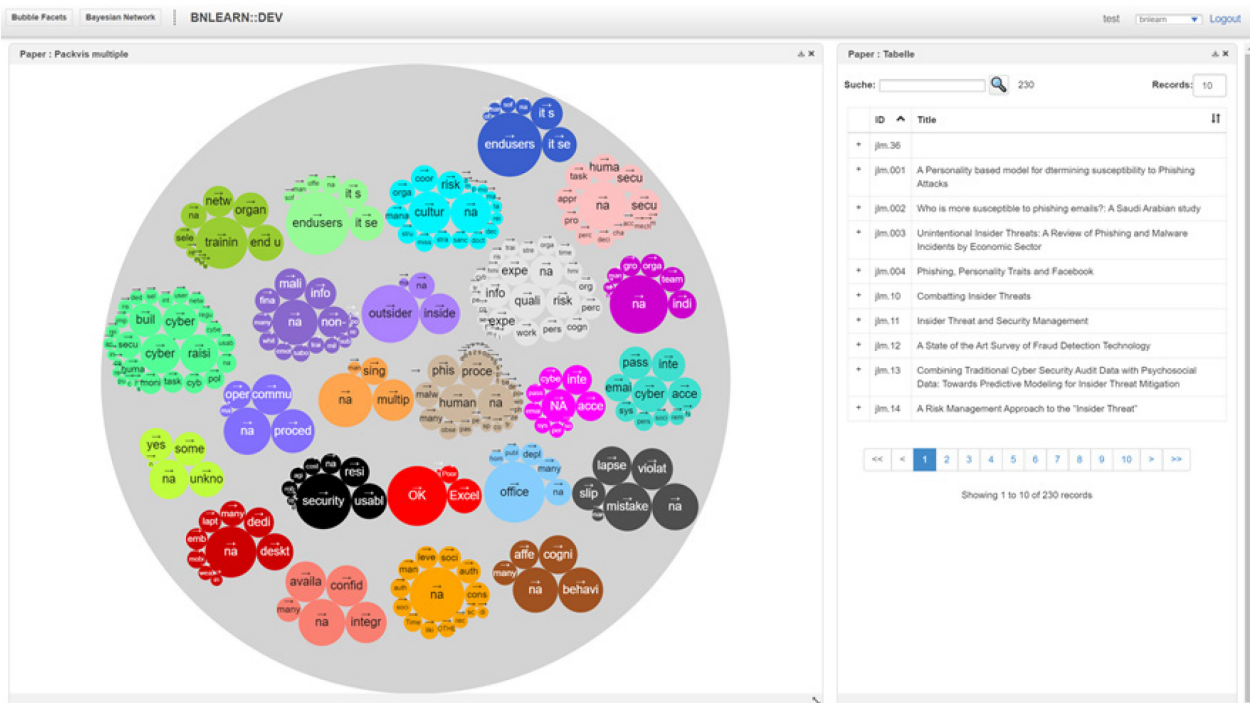


Figure 9-1: Dashboard with Widgets for Facet Exploration with Tag-Clouds in Circle Pack Layout.

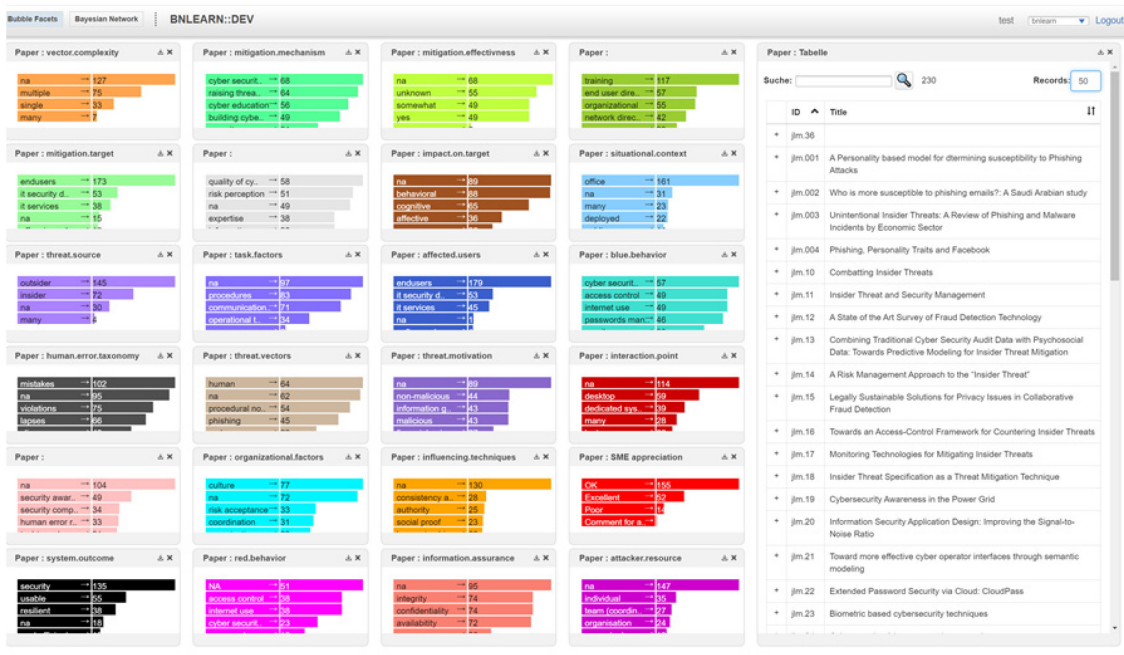


Figure 9-2: Dashboard with Widgets for Facet Exploration with Tag-Clouds in Horizontal Bar Layout.

The advantage of the circle pack layout is that each tag is shown, whereas in the horizontal bar layout some tags are hidden and the widget must be scrolled to bring them into view. However, the advantage of the horizontal layout is that the order top tags of each category and its labels can be read very easily. The tag-clouds can be used to select tags for filtering. The tags selected are then listed in a viewport for the filter. Here it can be toggled if all selected tags or at least one tag per category should occur in the filtered papers. In the visualizations the selected tags can also be inverted which allows papers to be filtered that are not containing this tag (Figure 9-3).

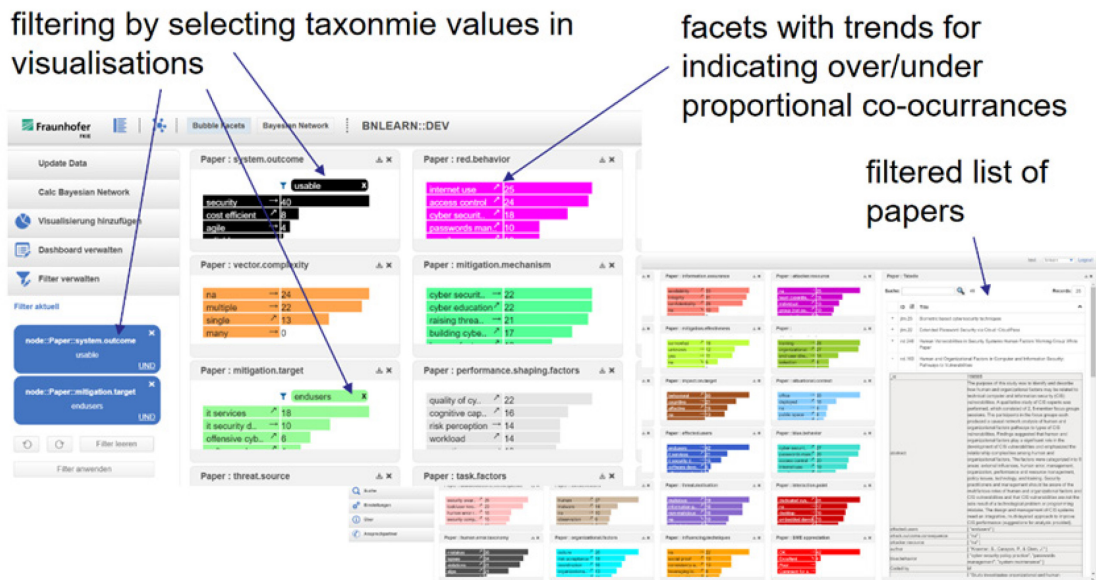


Figure 9-3: Facet Exploration by Elements of Taxonomy – Details.

The filter is immediately applied to each visualization. The tag-clouds adapt to the filter and show the new number of occurrences of tags in the filtered papers. To indicate if a tag occurs more or less in proportion to the overall population a small arrow is displayed at each tag. The table of papers also shows only the papers fitting the filter. The table facilitates the application of a full text search on the papers. The full text search influences the tag-clouds and the trends for each tag as well (Figure 9-4). This provides an effective means for users to explore the topics discussed in the papers.

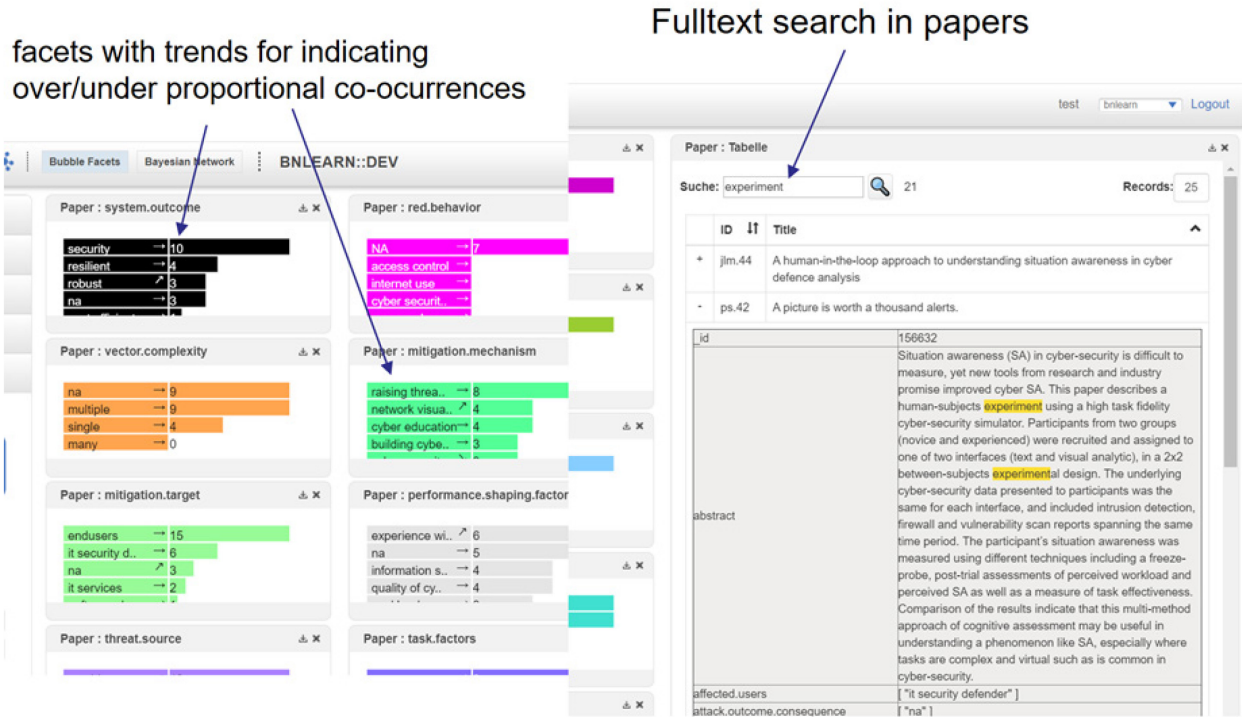


Figure 9-4: Facet Exploration by Elements of Taxonomy – Details.

9.5.2 Bayesian Network Exploration

The trends are of special interests, because they indicate if the occurrence of one tag in the database influences the co-occurrence with another tag. In fact, the analysis with algorithms for Bayesian network structure learning look into each combination of subsets of concepts in the joint probability distribution to assess if they are statistically dependent/independent.

We put the data into standard Bayesian network structure learning algorithms, wrote the Bayesian network into the graph database, and used the visualization framework to visualize and explore the Bayesian network. Figure 9-5 shows the result. On the left, each node represents a specific tag of the concepts. The concepts are color-coded. The network in the middle aggregated the tags of each concept into a single node.

The Bayesian network algorithm did not find links to all concepts, which indicated that these concepts should be considered as independent from the others, but this seems suspicious. To facilitate the inspection of the co-occurrences of tags between two concepts Sankey-diagrams has been used. By selecting one concept the Sankey shows their tags as source and targets and the co-occurrence as links between them. The size of the links scales with the number of co-occurrences. If a second concept is selected, the tags of the first selected concepts are displayed as sources and the tags of the second selected concepts as the targets in the Sankey. Figure 9-6 shows an example.

HFM-259 DATA EXPLORATION

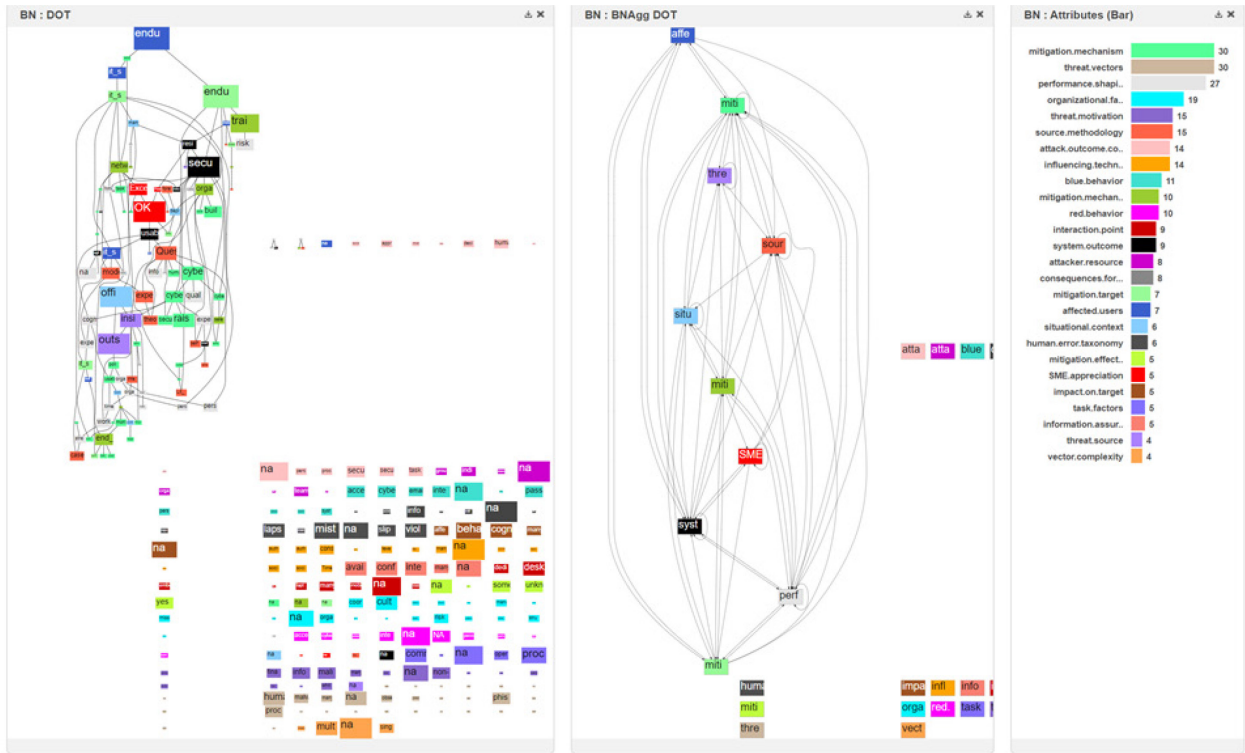


Figure 9-5: Bayesian Network for All Occurrences of All Values (Left) and Aggregated to Category-Elements (Middle) Bar Chart as Legend for Color Code of Category.

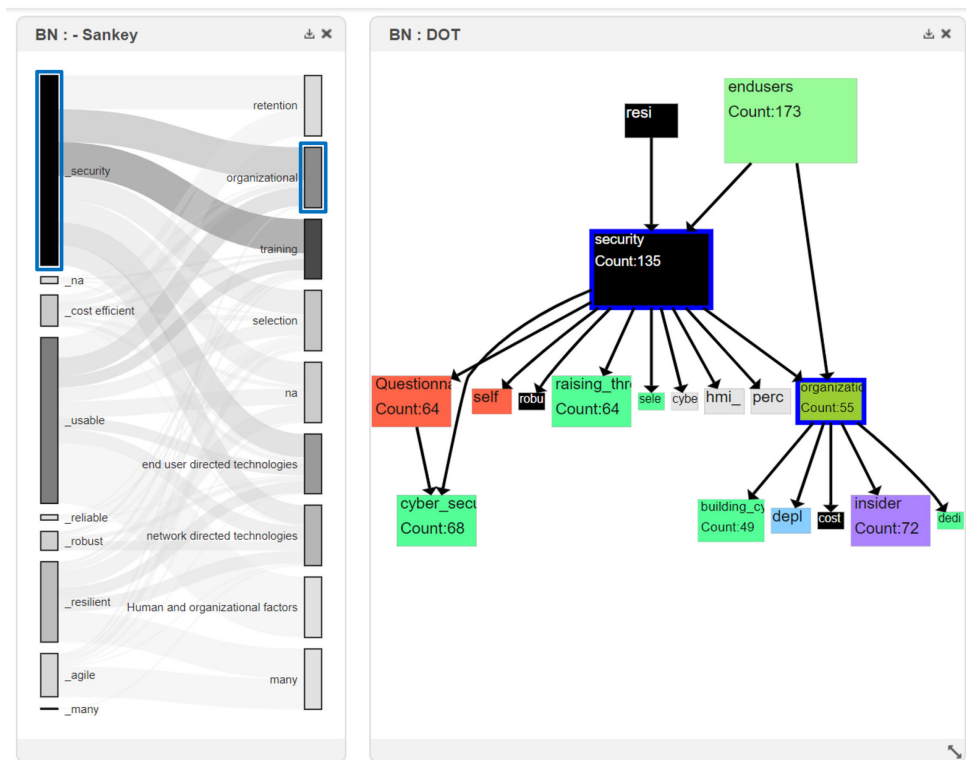


Figure 9-6: Selecting Nodes in Network Diagram to Show Distribution of Co-Occurrences in (Normalized) Sankey.

Usually, users are interested in how the distribution for one tag changes when it is filtered to another. If only the absolute numbers are displayed, the first thing that catches the eye are the connections that occur frequently. But this can also be the case if the tag in the target node of the Sankey is also common in the population. In order to overcome this, we apply a visualization approach in which the size of the target nodes are normalized, i.e., of equal size (Figure 9-7). Thus, connections of source nodes that are disproportionately frequent to a target node are more noticeable. However, this does not work so well for nodes that occur proportionally less often.

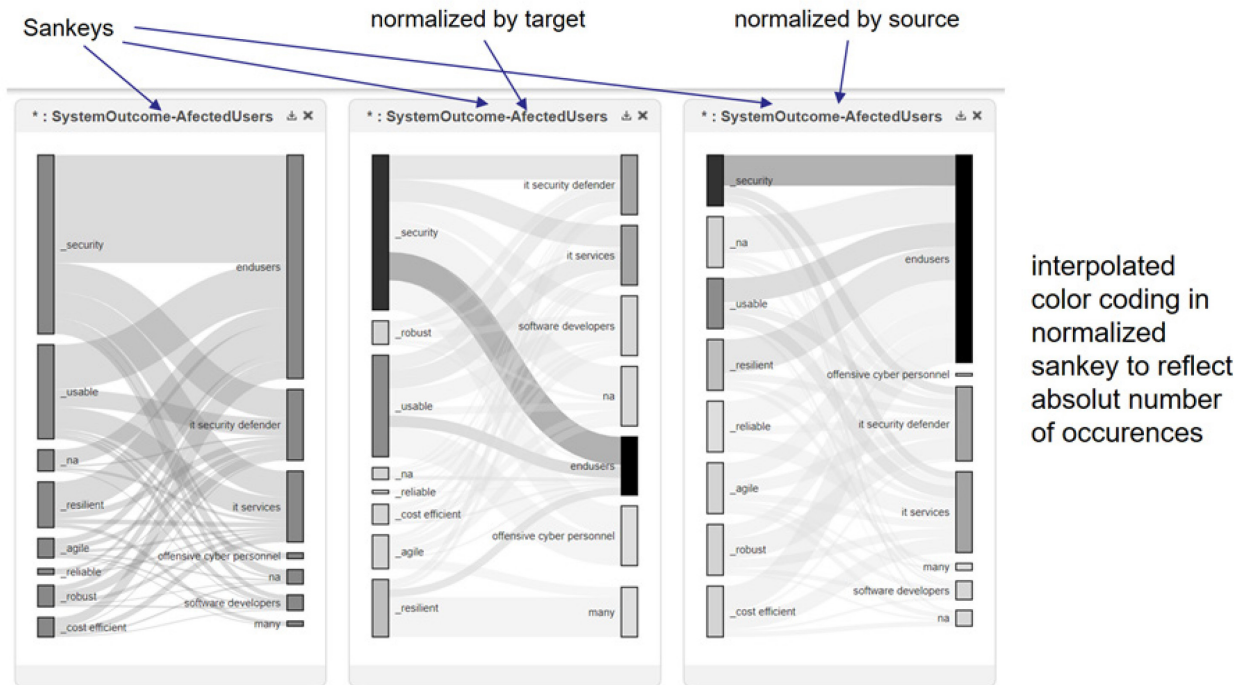


Figure 9-7: Normalized Sankeys.

9.6 CONCLUSION

In this chapter, we have presented an interactive dashboard to explore the ontology that was developed by Alice the HFM-259 RTG (Human Systems Integration Approach to Cyber Security) from 230 papers that discuss Human Factors in Cyber Security. The approach is very generic and shows how generally a document collection can be explored when they have been tagged with attributes of concepts of an ontology.

The goal was to identify possible relationships that result from the tags shared among the 230 documents. A central point was to analyze how the presence of one tag affects the presence of other tags. For this purpose, algorithms were used to determine the structure of a Bayesian network that represents the corresponding statistical relationships and this is integrated into the dashboard for visualization. One challenge was that there were many possible tags (288) over a relatively small number of documents, i.e., 230. Even though a lot of tags were assigned to each document, because at least one had to be selected from different categories, the Bayesian network was very extensive, but no clear structure could be recognized. The interactive analysis of the concrete distribution of the tags with the help of the Sankey diagram also raises concerns about the model, both when analyzing the links included in the Bayesian network and when looking at isolated tag-nodes.

It is recommended that future work should look to develop a new approach that can automatically derive corresponding relationships from the data. Word relationship networks could also be used, in which links between tags indicate the co-occurrence in a document. Although, at present, these networks contain too many links, which results in lack of clarity, this can be addressed by making a pre-selection based on an ontology of the concepts/categories of the tags, as it has been done in Ref. [5] for word relationship networks – though they do not consider or represent conditional probabilities.

9.7 REFERENCES

- [1] NATO STO (2020). Human Systems Integration Approach to Cyber Security. Final Report of Research Task Group HFM-259. STO-TR-HFM-259. NATO STO: Neuilly-sur-Seine, France.
- [2] Scutari, M. (2009). Learning Bayesian Networks with the bnlearn R Package. arXiv preprint arXiv:0908.3817.
- [3] Russell, S.J., and Norvig, P., (2009). Artificial Intelligence: A Modern Approach, 3rd edition. Prentice Hall.
- [4] Bostock, M., Ogievetsky, V., and Heer, J. (2011). D³ data-driven documents. IEEE Transactions on Visualization and Computer Graphics 17(12), pp. 2301-2309.
- [5] Vliet, V.T. (2020). Human Behaviour in Cyber Security: A Knowledge Base Perspective. Chapter 3 in NATO STO, Human Systems Integration Approach to Cyber Security. Final Report of Research Task Group HFM-259. STO-TR-HFM-259. NATO STO: Neuilly-sur-Seine, France.

Chapter 10 – CONCLUSIONS AND RECOMMENDATIONS

Dr. Margaret Varga
University of Oxford
UNITED KINGDOM

Dr. Elena Camosis
NATO STO CMRE
ITALY

Dr. Petter Bivall
Swedish Defence Research Agency
SWEDEN

Valérie Lavigne
Defence Research and Development
CANADA

10.1 CONCLUSIONS

The ever growing volumes of big, complex, often messy, and uncooperative data ubiquitously available cannot be exploited and understood using conventional approaches, thus alternative approaches are sought. The application of visualization and visual analytics together with human factor methods makes it possible to explore, analyze, understand, and exploit such challenging data. Scientific and technological research addressing defence and security domains such as: cyber security, maritime domain awareness, social media and in situ / post analysis of simulations have been undertaken by IST-141, and these have been discussed in this report. The effectiveness of the techniques has been demonstrated for providing acute situation awareness capabilities and understanding, and thus enhancing NATO's attainment in Information Superiority.

Interactive exploratory visual analytics and visualization techniques have been researched and developed by the Group's members to support interactive exploration, analysis and understanding of large and complex data. These techniques were applied to a cross-section of datasets such as Ukraine IED data (NATO C-IED COE), HFM-259 ontology data, simulation data, social media data and cyber network data. This work has demonstrated the high level of effectiveness of the approaches for knowledge discovery and gaining insight (situation awareness) to support informed decision making.

The Group has explored and developed the applications of visual analytics and visualization relevant to specific users as well as generic users (such as the general public for public engagement). The Group has stressed the importance of human factors and the understanding of mental models in meeting user needs.

Chapter 1 gives an overview of the work that the Group has undertaken.

Chapter 2 discusses the importance of human factors in visual analytics.

Chapter 3 reports the application of visual analytics and visualization in addressing the challenges in maritime domain awareness.

Chapter 4 and Chapter 5 are concerned with the exploration of social media and simulation data respectively.

Chapter 6 gives an overview of the exploration of visual analytics in deep learning.

Chapter 7 reports on the application of visual analytics for situation awareness in the cyber domain; this chapter also provides a discussion on symbology in the cyber domain.

The IED storytelling described in Chapter 8, provides an effective and intuitive means for a broad range of users, including the general public (public engagement), to gain insight about IED incidents.

Chapter 9 discusses the collaboration between IST-141 and HFM-259 in the exploration of the HFM-259 dataset relating to research papers on human factors in cyber security.

CONCLUSIONS AND RECOMMENDATIONS

IST-141 raised awareness of their visual analytics and visualization work. The Group facilitated the exploitation of visual analytics and visualization technologies, broadened the horizon of the understanding and exploration of these technologies within NATO and partner countries and beyond. The Group collaboration provided high leverage to the work of the individual participating nations through the generation and sharing of new ideas, tools, and data, as well as through the Group's continued development of its predecessor Group's inter-Panel / inter-Group workshop mechanism.

The Group organized, and / or participated in, and / or presented its work in, the following fora:

- IST-HFM-154 Specialists' Meeting: Cyber Symbolology, 28th – 30th November 2016, Dayton, USA.
- IST-178 Inter-Panel and Inter-Group Workshop: Big Data Challenges: Situation Awareness and Decision Support, 15th – 16th October 2019, Budapest, Hungary.
- IEEE VizSec, 2017, 2018 and 2019.
- IEEE VIS, 2016.
- International Conference on Human-Computer Interaction, 2018.
- NATO IST-143 Cyber Security Science and Engineering Lecture Series (2016 – 2019).
- NATO IST-170 Cyber Security Science and Engineering Lecture Series (2019 – present).

The Group had joint meetings with:

- HFM-259 at FKIE, Fraunhofer, Germany, on 17th October 2017.
- ENSC, IST-144 and IST-157, Bordeaux, France, 28th May 2018.
- SAS-139 at CMRE, Italy, 25th, 26th and 27th March 2019.
- IST-177 in Budapest, Hungary, 18th October 2019.

The Group also presented their work at:

- IST-129 RTG meeting in London, 3rd May 2016.
- HFM-294 RTG meeting in Toronto, Canada, 29th October 2018.

The Group has 32 publications.

These activities not only maintain awareness and continuity, but also provide invaluable opportunities to interact and develop new ideas and collaborations within and beyond NATO.

10.2 RECOMMENDATIONS

Interactions with, and support from, real users and the availability of real data are essential factors in developing and deploying effective visual analytics and visualization techniques and tools. Both issues require high-level support and leadership from within NATO and member nations, and this Group strongly recommends work be carried out to address this – e.g., to develop forms of modus operandi which are widely recognized (e.g., the IEEE VAST challenge datasets) outside the immediate S&T environment. The lack of such interaction represents a very real hinderance to leading edge advancement and deployment.

Visual analytics and visualization together with human factors, modelling and simulation and related analysis tools and techniques are key enabling technologies which provide one of the possible links to bring together disparate groups of researchers and users. The inter-Group inter-Panel workshop mechanism pioneered by IST-141 and its predecessor group IST-110 have proved to be an effective interaction enabler.

IST-141 thus recommend regular joint inter-Panel / inter-Group activities across the full spectrum of NATO S&T activities to bring members of different RTGs and Panels together to enable them to explore and benefit from each other's work, and to develop collaborations to complement and enhance their respective efforts, as well as facilitating the sharing of data, tasks, and tools.

The IST-141 Group members had hoped that such activities would attract 'real users' to interact with the researchers, and indeed, some users did participate, but further work and high-level support is required to build on this.



REPORT DOCUMENTATION PAGE			
1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	STO-TR-IST-141 AC/323(IST-141)TP/1079	ISBN 978-92-837-2396-7	PUBLIC RELEASE
5. Originator	Science and Technology Organization North Atlantic Treaty Organization BP 25, F-92201 Neuilly-sur-Seine Cedex, France		
6. Title	Exploratory Visual Analytics		
7. Presented at/Sponsored by	This is the Technical Report of the NATO IST-141 Research Task Group "Exploratory Visual Analytics."		
8. Author(s)/Editor(s)	Multiple	9. Date	February 2023
10. Author's/Editor's Address	Multiple	11. Pages	124
12. Distribution Statement	There are no restrictions on the distribution of this document. Information about the availability of this and other STO unclassified publications is given on the back cover.		
13. Keywords/Descriptors	Artificial intelligence; Batch processing; Counter Improvised Explosive Device (C-IED); Decision support; Deep learning; Exploratory visual analytics; Improvised Explosive Device (IED); Maritime situation awareness; Network analysis; Simulation analysis; Situation awareness; Social media; Storytelling; Trend analysis; Visualization		
14. Abstract	<p>Information superiority is one of the primary enablers for military dominance; the exploitation of all relevant information from multiple sources is a key factor for NATO's information superiority. Visualization and visual analytics research are essential to address the needs of the 2015 NATO targets of emphasis in Information Analysis (IA) and Decision Support (DS): IA&DS-1 on Decision Support and IA&DS-2 on Big Data and Long Data Processing and Analysis.</p> <p>Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. The Group investigated, researched and fostered collaborations in knowledge extraction and data analysis for timely situation awareness to support effective decision making. The IST-141 group researched, developed and applied exploratory visual analytics techniques: 1) To exploit and make sense of large and complex data sets, i.e., Big Data; 2) To help make tacit knowledge explicit; 3) To provide acute situation awareness, and 4) To support informed decision making across a wide range of defence and security application domains including cyber, maritime, genomics and social media domains, as well as post analysis and in situ visualization for simulation data.</p>		





BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cs.o.nato.int



DIFFUSION DES PUBLICATIONS
STO NON CLASSIFIEES

Les publications de l'AGARD, de la RTO et de la STO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la STO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la STO, de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre et la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la STO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (<http://www.sto.nato.int/>) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE

Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7, D-53229 Bonn

BELGIQUE

Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research
for Defence, National STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

BULGARIE

Ministry of Defence
Defence Institute "Prof. Tsvetan Lazarov"
"Tsvetan Lazarov" bul no.2
1592 Sofia

CANADA

DGSIST 2
Recherche et développement pour la défense Canada
60 Moodie Drive (7N-1-F20)
Ottawa, Ontario K1A 0K2

DANEMARK

Danish Acquisition and Logistics Organization
(DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESPAGNE

Área de Cooperación Internacional en I+D
SDGPLATIN (DGAM)
C/ Arturo Soria 289
28033 Madrid

ESTONIE

Estonian National Defence College
Centre for Applied Research
Riia str 12
Tartu 51013

ETATS-UNIS

Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72
92322 Châtillon Cedex

GRECE (Correspondant)

Defence Industry & Research General
Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE

Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25
H-1885 Budapest

ITALIE

Ten Col Renato NARO
Capo servizio Gestione della Conoscenza
F. Baracca Military Airport "Comparto A"
Via di Centocelle, 301
00175, Rome

LUXEMBOURG

Voir Belgique

NORVEGE

Norwegian Defence Research
Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE

Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL

Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE

Vojenský technický ústav s.p.
CZ Distribution Information Centre
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

ROUMANIE

Romanian National Distribution
Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353 Bucharest

ROYAUME-UNI

Dstl Records Centre
Rm G02, ISAT F, Building 5
Dstl Porton Down
Salisbury SP4 0JQ

SLOVAQUIE

Akadémia ozbrojených síl gen.
M.R. Štefánika, Distribučné a
informačné stredisko STO
Demänová 393
031 01 Liptovský Mikuláš 1

SLOVENIE

Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

TURQUIE

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi
Başkanlığı
06650 Bakanlıklar – Ankara

AGENCES DE VENTE

**The British Library Document
Supply Centre**
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

**Canada Institute for Scientific and
Technical Information (CISTI)**
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa, Ontario K1A 0S2
CANADA

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications STO, RTO et AGARD figurent dans le « NTIS Publications Database » (<http://www.ntis.gov/>).



BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cs.o.nato.int



**DISTRIBUTION OF UNCLASSIFIED
STO PUBLICATIONS**

AGARD, RTO & STO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all STO reports, or just those relating to one or more specific STO Panels, they may be willing to include you (or your Organisation) in their distribution.

STO, RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of STO reports as they are published, please visit our website (<http://www.sto.nato.int/>) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM

Royal High Institute for Defence –
KHID/IRSD/RHID
Management of Scientific & Technological
Research for Defence, National STO
Coordinator
Royal Military Academy – Campus
Renaissance
Renaissancelaan 30
1000 Brussels

BULGARIA

Ministry of Defence
Defence Institute “Prof. Tsvetan Lazarov”
“Tsvetan Lazarov” bul no.2
1592 Sofia

CANADA

DSTKIM 2
Defence Research and Development Canada
60 Moodie Drive (7N-1-F20)
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC

Vojenský technický ústav s.p.
CZ Distribution Information Centre
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

DENMARK

Danish Acquisition and Logistics Organization
(DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESTONIA

Estonian National Defence College
Centre for Applied Research
Riia str 12
Tartu 51013

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc – BP 72
92322 Châtillon Cedex

GERMANY

Streitkräfteamt / Abteilung III
Fachinformationszentrum der
Bundeswehr (FIZBw)
Gorch-Fock-Straße 7
D-53229 Bonn

GREECE (Point of Contact)

Defence Industry & Research General
Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY

Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25
H-1885 Budapest

ITALY

Ten Col Renato NARO
Capo servizio Gestione della Conoscenza
F. Baracca Military Airport “Comparto A”
Via di Centocelle, 301
00175, Rome

LUXEMBOURG

See Belgium

NETHERLANDS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY

Norwegian Defence Research
Establishment, Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND

Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL

Estado Maior da Força Aérea
S DFA – Centro de Documentação
Alfragide
P-2720 Amadora

ROMANIA

Romanian National Distribution Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353 Bucharest

SLOVAKIA

Akadémia ozbrojených síl gen
M.R. Štefánika, Distribučné a
informačné stredisko STO
Demänová 393
031 01 Liptovský Mikuláš 1

SLOVENIA

Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

SPAIN

Área de Cooperación Internacional en I+D
SDGPLATIN (DGAM)
C/ Arturo Soria 289
28033 Madrid

TURKEY

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanlıklar – Ankara

UNITED KINGDOM

Dstl Records Centre
Rm G02, ISAT F, Building 5
Dstl Porton Down, Salisbury SP4 0JQ

UNITED STATES

Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

SALES AGENCIES

**The British Library Document
Supply Centre**
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

**Canada Institute for Scientific and
Technical Information (CISTI)**
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa, Ontario K1A 0S2
CANADA

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of STO, RTO and AGARD publications are given in “NTIS Publications Database” (<http://www.ntis.gov>).