

**Project Number:** 898588

**Project Acronym:** VISIONS

**Project title:** Neural Video Processing and Streaming for Real-time Traffic Monitoring

## **DATA MANAGEMENT PLAN**

# Data Management Plan

Beneficiary: University of Exeter  
Department: Department of Computer Science  
Name of Fellow: Dr. Xu Zhang  
Name of Supervisor: Prof. Geyong Min

## 1. Data Summary

This project participates in the Horizon 2020 Pilot on Open Research Data. The provisions for open access to research data are described in Grant Agreement Arts 29.3 *Open access to research data* and ARTICLE 39 *Processing of personal data*. This document follows the template for the H2020 Data Management Plan v1.0 (13/10/2016), and will be regularly reviewed and updated as necessary.

### 1.1. Objectives of the VISIONS project

The overarching aim of VISIONS is to enable the real-time traffic monitoring in smart cities, through leveraging the emerging machine learning based methods in video processing and streaming to reduce the required network bandwidth and guarantee the Quality of Experience (QoE) perceived by users in the dynamic network. To achieve the general aim, several specific objectives have been designed, spanning the field of video processing and networking:

- ❖ Obj1: Allocating different qualities to different macroblocks within the traffic monitoring video to reduce the required bandwidth without noticeable visual quality degradation, which conforms to the feature of human visual system and the network bandwidth constraints.
- ❖ Obj2: Enhancing video quality through using DNN, in order to increase the user concurrency in the cloud and reducing the streaming bandwidth over the network simultaneously.
- ❖ Obj3: Adapting the video bitrate by using reinforcement learning based method, in order to accommodate the unexpected network dynamics, guaranteeing users' QoE.
- ❖ Obj4: Designing a novel real-time traffic monitoring framework, including the above three techniques, and verifying the overall framework in the real application scenario.

### 1.2. Types of data utilized and generated

To achieve the above ambitious aim and evaluate the proposed approaches, VISIONS will both collect existing relevant data from the open-source datasets on the Internet, and generate new data within the project. On the one hand, all existing datasets will be used in accordance with any licences or access requirements, and to avoid concerns of ownership and copyright, these datasets will not be archived. Instead, the link to the original data will be added as references in the relevant publications. On the other hand, the generated data will not contain any personal privacy information to protect users' privacy. To this end, the evaluation of VISIONS will be carried out in a pure experimental environment. In the following, the type, source, format, size of the data and the link to which objective will be elaborated in Table 1.

Table 1. Types of data utilized and generated in VISIONS

Public Dataset					
Data	Description	Source	File format	Objective	Size
BVI-DVC	BVI-DVC is a public large video dataset for training DL-based video compression systems, with specific emphasis on machine learning tools that enhance conventional coding architectures, including spatial resolution and bit depth up-sampling, post-processing and in-loop filtering. BVI-DVC contains 800 sequences at various spatial resolutions from 270p to 2160p and has been evaluated on ten existing network architectures for four different coding tools.	BVI-DVC: A Training Database for Deep Video Compression ( <a href="https://vilab.blogs.bristol.ac.uk/2020/02/2375/">https://vilab.blogs.bristol.ac.uk/2020/02/2375/</a> )	.mp4	Obj1, Obj2	396G
FCC broadband datasets	The datasets are collected by the Measuring Broadband America (MBA) program of FCC (Federal Communication Commission). The program began in 2010 and has since then provided periodic reports on a sample set of consumers served by Internet Service Providers (ISPs) who serve approximately eighty-five percent of broadband consumers in the United States. The FCC only uses data gathered in agreed-upon reference months and audited according to the established methodologies in the comparative analysis of ISPs included in its reports.	Raw Data - Measuring Broadband America 2016 ( <a href="https://www.fcc.gov/reports-research/reports/measuring-broadband-america/raw-data-measuring-broadband-america-2016">https://www.fcc.gov/reports-research/reports/measuring-broadband-america/raw-data-measuring-broadband-america-2016</a> )	.gz	Obj3, Obj4	42.4G
Glas video streaming performance datasets	The datasets were collected by University of Glasgow in 2009 and 2010. They contain measurements of CBR RTP traffic, sent from a campus machine to receivers connected to the Internet via ADSL and Cable links. The processed datasets contain end-to-end queueing delay time series, and end-to-middle delay measurements and packet-pair capacity measurements.	Measurements of Real-Time Traffic to Residential Users ( <a href="http://web.cs.wpi.edu/~claypool/mmsys-dataset/2011/isp/author.html">http://web.cs.wpi.edu/~claypool/mmsys-dataset/2011/isp/author.html</a> )	.gz	Obj3, Obj4	2.6G
Generated Dataset					
Data	Description	File format	Objective	Estimated Size	

Synthetic network conditions	To improve the robustness of the DRL model for video bitrate adaption, we will synthesize some extra network traces to train the DRL model. The data will include the timestamp, the IPs of the source and destination, the network delay, the available bandwidth.	.logs	Obj3	200M
Performance dataset	To evaluate the proposed approaches in VISIONS, we will carry out the experiments in a pure experimental environment and collect the performance data from the viewing devices. The data will include the timestamp, the IPs of the source and destination, the delay perceived by users, the available bandwidth, the buffering time, the buffer occupancy, and the video bitrate.	.logs	Obj4	100M

### 1.3. Data utility

The data generated by VISIONS will be made open access to the public. It can be utilized by the broader scientific community to reproduce the experiments presented in the publications of VISIONS. Furthermore, the data can also be leveraged by researchers in the field of video streaming and traffic monitoring to develop more advanced algorithms to further enhance the QoE perceived by users and improve the resource efficiency in ITS (Intelligent Transport System) in the future.

## 2. FAIR data

VISIONS complies with the guiding principles for scientific data management outlined by the H2020 programme, guaranteeing that the research data is Findable, Accessible, Interoperable and Reusable (FAIR).

### 2. 1. Making data findable, including provisions for metadata

#### Persistent Digital Object Identifier (DOI)

The datasets generated by VISIONS will be archived in the institutional repository Open Research Exeter (ORE), which is maintained by the beneficiary, University of Exeter, to support the principles of open access to research publications and research data. ORE is fully searchable and indexed by search engines to ensure that the data are highly discoverable. Note that a unique DOI will be assigned to the dataset once it is uploaded to the data repository, which will also be linked to appropriate publication records in ORE. With the DOI, the dataset can be accessed and cited in research papers, which can further enhance the visibility of datasets.

#### Metadata

Each dataset deposits in ORE will be accompanied by Dublin Core metadata to describe the information regarding the owner, date accessioned, date issued, access conditions, generation methods, time references, structure and organisation of data files, file formats, variable names, labels and descriptions of variables and values. Moreover, the terms “Marie Skłodowska-Curie Actions”, the name of the action (H2020-MSCA-IF-2019), acronym (VISIONS) and grant number (898588) will be included, along with the DOI and the publication date. Upon archiving the dataset, key

words will be provided to enhance the findability of the data, thus optimizing the possibility for re-use by other researchers in the related field.

### **Naming conventions**

VISIONS datasets will be named with project title (VISIONS), the No. of WP to which the data is linked, the purpose of the data, and the version. Specifically, the naming rule is as follows:

VISIONS -WPx-purpose-vx.

For example, the dataset entitled with “VISIONS-WP01-QualityAllocation-v1” is the dataset with an initial version (v1) for video quality allocation in WP01.

The naming convention will be explained in documentation (e.g., a README file) that will accompany the data deposited to ORE so that others will be able to understand the data, increasing the re-use of data.

## **2.2. Making data openly accessible**

All the data generated from VISIONS, including the metadata documentation and code, will be deposited in the ORE on publication of the research papers, which can be utilized to validate the results presented in the deposited scientific publications. The ORE is the University of Exeter’s online repository for storing and providing access to research output, which can ensure that the research data of VISIONS is securely preserved in a long-term manner. ORE can increase the visibility of publications and research data by making them available as open access and enabling compliance with open access policies of the University and funders of research. The data will be openly accessible to everyone, with the metadata being publicly searchable and discoverable, which will indicate how and on what terms the dataset can be accessed.

The existing datasets will not be deposited to ORE, but the documentation accompanying the data in ORE will include links to these existing datasets so that anyone who downloads your data from ORE will be able to replicate the analysis.

## **2.3. Making data interoperable**

### **Standards and formats**

The research data generated from VISIONS will be made publicly accessible, where legally, commercially, and ethically appropriate, to the research community. To make it interoperable, it will be archived in standard formats and in compliance with available open software applications to facilitate data exchange and reuse. The file formats for different kinds of data are listed in Table 1. For better interoperability of the data, a README file will be created, which will explain the data in terms of variable names, labels and descriptions of variables and values, the file naming convention, links to the existing datasets, any tools/software required to access/use the data etc. For instance, MATLAB or Python may be needed to analyze the research data.

## **2.4. Increase data re-use (through clarifying licences)**

### **Licencing**

The data deposited in the University’s institutional repository ORE will be made available under ORE policy, which allows equivalent rights to the Creative Commons Attribution (CC BY) licence to all items in the repository, unless otherwise specified. It can maximise the re-use of data while still ensuring the credit be given to the dataset creator. Upon the publication of a research paper, the related data will be deposited and made available for third parties, which are retained essentially indefinitely and

could be accessed by others after the end of the project (31/7/2023).

### **3. Allocation of resources**

#### **Costs**

For the cost of data collection and data processing, the fellow will discuss with the beneficiary (supervisor and research manager/financial officer), the Coordinator and Project Manager about the budget, who will ensure the VISIONS Grant Agreement and this DMP are followed.

Costs related to data management are expected to be limited and will be covered by the beneficiary's institutional unit costs. Besides, after archiving in ORE, the cost for the long-term preservation will also be covered by the beneficiary itself.

For the costs related to the administrative coordination of data management and deposition in line with this DMP, it will be covered by project funds at network level administered by the coordinator.

#### **Responsibilities**

The coordinator, Prof. Geyong Min, is responsible for implementing the data management of VISIONS, while the fellow, Dr. Xu Zhang, is responsible for the data generation, metadata production, data quality guaranteeing during the whole process of the project. Upon the publication of a paper, the fellow should also be in charge of uploading the related data to ORE, and backing up it in the server provided by the High Performance Computing and Networking (HPCN) Research Group led by Prof. Min. The Open Research team from the University of Exeter will be responsible for the management of ORE.

### **4. Data security**

The beneficiary will provide the fellow several options for secure storage of data. During the data collection and experimental phase, apart from his personal computer, the fellow will back up the data in the password-protected servers offered by the host research group, the High Performance Computing and Networking (HPCN) Research Group led by Prof. Min, in the College of Engineering, Mathematics and Physical Sciences, University of Exeter.

At the end of the project, the data will be deposited for long-term storage in the institutional repository, the Exeter ORE data archive, which is maintained by the hosting beneficiary, University of Exeter.

### **5. Ethical aspects**

There are no ethical issues concerning VISIONS project.

### **6. Other issues**

No, VISIONS does not use other national/funder/sectorial/departmental procedures for data management.

## **7. Further support in developing your DMP**

The dataset within VISIONS will be stored at the ORE, which is maintained by the hosting beneficiary, University of Exeter. No other support will be needed in developing the DMP.

HISTORY OF CHANGES		
Version	Publication date	Change
1.0	31.01.2022	▪ Initial version