Plan Overview

A Data Management Plan created using DMPonline

Title: "Intelligent Operational Reliability for Automotive Systems" (UoB PhD 21MR)

Creator: Alexey Uglanov

Principal Investigator: Felician Campean

Data Manager: Alexey Uglanov

Project Administrator: Felician Campean

Contributor: Alexey Uglanov

Affiliation: University of Bradford

Template: DCC Template

ORCID iD: 0000-0003-4166-8077

Project abstract:

This project, carried out in conjunction with a major global automotive manufacturer, aims to develop and validate machine learning algorithms for semantic processing of heterogeneous data available from operational automotive systems warranty databases, and the automatic generation of data models. This research will consider challenges with machine learning in multidimensional heterogeneous data sets (both text and variable data) stemming from the multiple types of data, accuracy and imbalance in the real world operational reliability data sets, which affects the robustness of the machine learning models. The validation of the methods and algorithms developed will be carried out in conjunction with real world data available from the automotive Company partner. The project will be based in the interdisciplinary Advanced Automotive Analytics Research Laboratory, part of the University of Bradford Automotive Research Centre, which has a strong track record of collaborations with the global automotive industry spanning over 25 years.

ID: 150695

End date: 31-05-2026

Last modified: 01-07-2024

Copyright information:

"Intelligent Operational Reliability for Automotive Systems" (UoB PhD 21MR)

Data Collection

What data will you collect or create?

ALL DATA

The collected research data includes heterogeneous data sources such as warranty claims data from the industrial partner, open-sourced passenger vehicle recall data from the US government and social networks (forum data). Most data is in tabular format stored or imported in CSV format, the forum data presumably in HTML format. The data collected is stored using a university-managed infrastructure (OneDrive and Teams), which will provide the necessary data security. The data coming from the industrial partner is stored in their internal databases. Backup and file recovery policies from both the University of Bradford and industrial partner are in place. Any data which are sensitive and confidential will be removed from public sharing which is in line with the data protection regulations within the United Kingdom. Sharing and long-term access will be available in CSV format, where possible.

WARRANTY DATA

- Description: the warranty claim dataset from the industrial partner, covering the vehicle model, components affected and other statistical data (dates, MiS, distances, textual fields with verbatim, etc.)
- Format: CSV for exported data, direct access to relational/vertical databases
- Volume: below 1 GB

RECALL DATA

- Description: NHTSA (National Highway Traffic Safety Administration) is a part of the Department of Transportation of the U.S. federal government responsible for transportation safety in the U.S.A. It covers safety standards writing and enforcing, administers the Vehicle Identification Number (VIN) system, defects investigations that lead to recalls issuance in case of safety risks. The recall database is open to the public (according to the Terms of Use) and helps both the customers (individual drivers) and the manufacturers (e.g., vehicle producers or component suppliers) to track any concerning issues. It was released at the end of 2002 and has been updated daily since then. NHTSA aims to cover the recalls from 1949 to present time, providing the details of the recall such as the vehicle model and component details, dates, chronology, remedy, defect summary, manufacturer-related information, and more. The full list of the TXT file fields is available here: https://static.nhtsa.gov/odi/ffdd/rcl/RCL.txt
- Format: ZIP archive with a tab-separated TXT file inside.
- Volume: ZIP archive 13 MB, tab-separated TXT file ~300 MB

FORUM DATA

- Description: the threads from the selected automotive forums
- Format: to be confirmed, presumably ZIP archive with HTML / CSV files
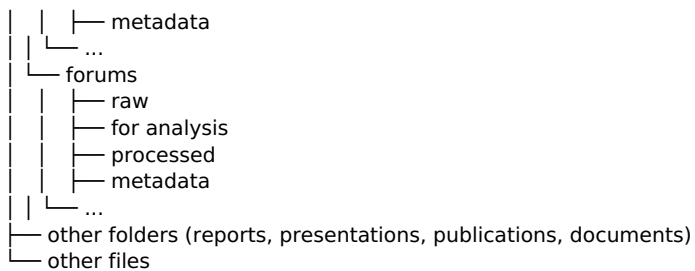- Volume: currently unknown

How will the data be collected or created?

ALL DATA

A standardised format for data collection, with agreed file naming conventions and version control (including the range of dates where applicable), is used by the investigators with guidance from the
University of Bradford's research data management team. Data type-specific sub-folders with further separation into subfolders of RAW / FOR ANALYSIS / PROCESSED are used for ease of data retrieval and usage.
Example:

```
project files
├── data
│   ├── warranty
│   │   ├── raw
│   │   │   └── warranty data model X component Y from date Z
│   │   ├── for analysis
│   │   ├── processed
│   │   ├── metadata
│   │   └── ...
│   ├── recalls
│   │   ├── raw
│   │   ├── for analysis
│   │   ├── processed
```

```
│   │   ├── metadata
│   │   └── ...
│   └── forums
│   │   ├── raw
│   │   ├── for analysis
│   │   ├── processed
│   │   ├── metadata
│   │   └── ...
├── other folders (reports, presentations, publications, documents)
└── other files
```

WARRANTY DATA

- Source: the industrial partner provides the warranty records created within the company.

RECALL DATA

- Source: public access files from NHTSA.

FORUM DATA

- Source: the automotive forum websites where the academic use is permitted / the owners of these websites where agreed

Documentation and Metadata

What documentation and metadata will accompany the data?

According to the "Research Data Management Policy" of the University of Bradford:

- "5.2.5 Research data should be accompanied by metadata that adheres to appropriate metadata standards."
- "5.2.6 Research data should also be identifiable, and retrievable when necessary. We encourage the use of persistent identifiers (where available) for datasets, for example the Digital Object Identifier (DOI), as well as ORCID for researchers."

ALL DATA

- Documentation will accompany the data in the form of descriptions with diagrams, where necessary, in MS Word and PDF formats. In addition to that, the data is described in the academic reports and reports for the industrial partner, as well as in the relevant conference/journal articles and final thesis.

WARRANTY DATA

- The metadata is shared by the industrial partner in CSV format with comments on each of the fields

RECALL DATA

- The metadata is available publicly on the NHTSA website in TXT format: https://static.nhtsa.gov/odi/ffdd/rcl/RCL.txt

FORUM DATA

- The metadata standards used depend on a particular website, from XML files to no standards at all. The creation of the CSV files about the parsed data is considered.

Ethics and Legal Compliance

How will you manage any ethical issues?

According to the "Research Data Management Policy" of the University of Bradford:

- "5.2.1. Researchers must comply with all legal, ethical, funding body and organisational requirements for the collection, use, re-use and storage of data, especially personal data, where particular attention should be paid to the requirements of the Data Protection Act 2018 and UK's General Data Protection Regulation (UK GDPR)."
- "5.2.3. They should maintain confidentiality where undertakings have been made to third parties or to protect intellectual property rights. Further expectations are elaborated in sections 6.7 and 6.8 of the University of Bradford Code of Practice for

[Research](#)"

Even though this project does not require ethical approval as there is no sensitive data gathered directly from any kind of participants (i.e., self-organised survey members) or the NHS is involved in the research, the ethics approval has been granted by the Chair of the Humanities, Social and Health Sciences Research Ethics Panel at the University of Bradford on 31/07/23. The Ethics Checklist number is EC27860.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

According to the "Research Data Management Policy" of the University of Bradford:

- "5.2.1. Researchers must comply with all legal, ethical, funding body and organisational requirements for the collection, use, re-use and storage of data, especially personal data, where particular attention should be paid to the requirements of the Data Protection Act 2018 and UK's General Data Protection Regulation (UK GDPR)."
- "5.2.2. The ownership of research data generated by persons employed by the University of Bradford is subject to the [Intellectual Property Policy](#). Researchers and their institutions should apply appropriate licences to clarify ownership and use of data."
- "5.2.3. They should maintain confidentiality where undertakings have been made to third parties or to protect intellectual property rights. Further expectations are elaborated in sections 6.7 and 6.8 of the [University of Bradford Code of Practice for Research](#)"

According to the "Intellectual Property Policy" of the University of Bradford:

- "1. The University through the provision of the Copyright Designs and Patents Act 1988, claims ownership of all intellectual property as specified in section (5), which is devised, made or created:
    - (a) by persons employed by the University in the course of their employment;
    - (b) by those engaged in study or research at the University who have agreed that this condition shall apply to them;
    - (c) by certain persons, under prior agreement, engaged by the University under contract for services during the course of or incidentally to that engagement."
- "5. The intellectual property of which ownership is claimed under (1) above includes:
    - (a) patentable and non-patentable inventions;
    - (b) works generated by computer hardware or software owned or operated by the University;
    - (c) videos, multimedia works, films, sound recordings, typographical arrangements and other works created with the aid of University facilities;
    - (d) registered and unregistered designs, plant varieties, genetic modifications and topographies;
    - (e) any works not included in (3) (a), (b), (c) or (d) commissioned by the University;
    - (f) databases, spreadsheets, firmware, courseware and related material not within (a), (b), (c), (d) or (e), if the University considers it to possess commercial potential;
    - and
    - (g) know-how and information associated with any of the above."
- "6. The University undertakes not to assert any claim to the ownership of intellectual property rights in: books, articles, plays, lyrics, scores, artistic works, other than ones commissioned by the University unless the work is deemed important to the protection and exploitation of other IP generated by these individuals in the course of their employment."

The contract and the NDA signed with the industrial partner cover the legal issues (Confidential Information disclosure and Confidentiality Notices, Intellectual Property rights and the University's Intellectual Property Policy, know-how, Financial Contributions, Limitations of Liability, Insurance, Force Majeure, and more) approved both by the company and the University of Bradford to ensure honest, safe, and reliable research.

WARRANTY DATA

- The industrial partner owns the data
- The sharing of data is restricted due to the GDPR and other regulations

RECALL DATA

- The public owns the data
- The data is publicly available and allowed for reuse: https://www.nhtsa.gov/nhtsa-datasets-and-apis

FORUM DATA

- The website owners own the data
- It is publicly available on the respective websites, however, there might be limitations on the sharing of the exported data and processed data

Storage and Backup

How will the data be stored and backed up during the research?

According to the "Research Data Management Policy" of the University of Bradford:

- "5.2.10 The University of Bradford will provide infrastructure and support for the safe storage of research data, as well as training for researchers at all levels that covers data security in line with current legislation, and data management planning."

According to the "Information Security Policy" of the University of Bradford:

- "University data (information) will be classified and provided with appropriate safeguards ensuring it is protected against unauthorised or inappropriate access or use and against accidental loss, destruction or damage thereby ensuring its confidentiality, integrity and availability."

The data collected is stored using a university-managed infrastructure (OneDrive and Teams), which will provide the necessary data security. The data coming from the industrial partner is stored in their internal databases. Backup and file recovery policies from both the University of Bradford and industrial partner are in place.
No additional charges for data storage are required at this stage.

How will you manage access and security?

According to the "Research Data Management Policy" of the University of Bradford:

- "5.2.10 The University of Bradford will provide infrastructure and support for the safe storage of research data, as well as training for researchers at all levels that covers data security in line with current legislation, and data management planning."

According to the "Information Security Policy" of the University of Bradford:

- "Applicable Statutory, Legal or National Best Practice Requirements:
  - General Data Protection Regulations (GDPR)
  - Computer Misuse Act 1990
  - ISO 27001:2013 Information Security Management Standard"
- "Current legislation which has a bearing on Information Security, includes but is not limited to:
  - Computer Misuse Act 1990
  - Copyright Designs and Patents Act 1988
  - Counter Terrorism and Security Act 2015 (in particular the 'prevent duty')
  - Data Protection Act 2018
  - Freedom of Information Act 2000
  - Privacy and Electronic Communications (EC Directive) Regulations 2003 (as amended)
  - Regulation of Investigatory Powers Act 2000"
- "University data (information) will be classified and provided with appropriate safeguards ensuring it is protected against unauthorised or inappropriate access or use and against accidental loss, destruction or damage thereby ensuring its confidentiality, integrity and availability."

The data collected is stored using a university-managed infrastructure (OneDrive and Teams), which will provide the necessary data security. The data coming from the industrial partner is stored in their internal databases. Access is given by the Principal Investigator only to the researchers involved in the project. The University password policy is followed, as well.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

According to the "Research Data Management Policy" of the University of Bradford:

- "5.2.7 Research data and the administrative records accompanying research projects should also be archived. Research data is a legitimate product of research and must be cited as such adhering to scholarly norms; re-used data must be explicitly traceable and original sources acknowledged."

All relevant research data will be preserved. A minimum of 10 years from the last data access request is a standard required term for storing the research data after the end of the project. Data relevant to journal articles will be preserved and shared,

What is the long-term preservation plan for the dataset?

According to the "Research Data Management Policy" of the University of Bradford:

- "5.2.11 In most cases research data and documentation relating to the development and administration of research projects should be kept for a minimum of ten years after the end of the project." Retention requirements are detailed in the Document Retention and Disposal Policy.

Data archiving is provided by the University via OneDrive.

Data Sharing

How will you share the data?

According to the "Research Data Management Policy" of the University of Bradford:

- "5.2.4 Research data should be accurate, complete, authentic and reliable, and made available in a timely way with as few restrictions as possible."
- "5.2.6 Research data should also be identifiable, and retrievable when necessary. We encourage the use of persistent identifiers (where available) for datasets, for example the Digital Object Identifier (DOI), as well as ORCID for researchers."

Where possible, the data will be made available in the publications. The data-sharing limitations are covered in the contract and NDA between the University and the industrial partner.

Are any restrictions on data sharing required?

Where possible, the data will be made available in the publications. The data-sharing limitations are covered in the contract and NDA between the University and the industrial partner.

Responsibilities and Resources

Who will be responsible for data management?

According to the "Research Data Management Policy" of the University of Bradford:

- "4.1.2. Primary responsibility for research data management lies with the **Principal Investigator**..."
- "4.1.3. When University staff participate in collaborative projects, a University of Bradford lead must be identified to take responsibility for the management of data produced by the University of Bradford."
- "4.1.4. **Principal Investigators (PI), Project Leads, and supervisors** (including Data Stewards) take responsibility to:
  - 4.1.4.1. Manage research data in accordance with the requirements of this policy.
  - 4.1.4.2. Ensure that they and the researchers under their supervision receive appropriate training in all aspects of research data management.
  - 4.1.4.3. Produce, adhere and update a DMP to address university, legal and funder requirements for the storage and retention of data.
  - 4.1.4.4. Ensure that data is accessible by appropriate authorised persons to maintain access.
  - 4.1.4.5. Produce metadata and documentation to describe data, to understand what research data exists, why, when and how it was generated and access restrictions and mechanisms.
  - 4.1.4.6. On completion of the research, ensure all relevant data are archived, deposited or disposed or appropriately, securely and auditability in line with the Research Data Management IT Security Guidance.
  - 4.1.4.7. Ensure if they leave the institution an accessible copy of the data produced under the auspices of the University is deposited before their departure.
  - 4.1.4.8. Ensure that all requirements imposed by funding bodies, regulatory agencies, third party data providers and collaborating partners are met"

What resources will you require to deliver your plan?

The already available resources provided by the University and the industrial partner are enough.

Planned Research Outputs

Book chapter - "An NLP-based framework for early identification of design reliability issues from heterogeneous automotive lifecycle data"

Natural Language Processing is increasingly used in different areas of design and product development with varied objectives, from enhancing productivity to embedding resilience into systems. In this paper, we introduce a framework that draws on NLP algorithms and expert knowledge for the automotive engineering domain, to extract actionable insight for system reliability improvement from data available from the operational phase of the system. Specifically, we are looking at the systematic exploration and exploitation of automotive heterogeneous data sources, including both closed-source (such as warranty records) and open-source (e.g., social networks, chatrooms, recall systems) data, to extract and classify information about faults, with predictive capability for early detection of issues. We present a preliminary NLP-based framework for enhancing system knowledge representation to increase the effectiveness and robustness of information extraction from data, and discuss the temporal alignment of data sources and insight to improve prediction ability. We demonstrate the effectiveness of the proposed framework using real-world automotive data in a recall study for a vehicle lighting system and a particular manufacturer: four recall campaigns were identified leading to corrective actions by the warranty experts

---

Planned research output details

Title	DOI	Type	Release date	Access level	Repository(ies)	File size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
An NLP-based framework for early identification of ...		Book chapter	Unspecified	Open	None specified	3 MB	Creative Commons Attribution Non Commercial No Derivatives 4.0 International	None specified	No	No