Communication in Emergency Management through Data Integration and Trust: an introduction to the CEM-DIT system

Fiona McNeill*
Heriot-Watt University, Scotland
f.mcneill@hw.ac.uk

Diana Bental
Heriot-Watt University, Scotland
d.s.bental@hw.ac.uk

Paolo Missier
Newcastle University
paolo.missier@ncl.ac.uk

Janetta Steyn
Newcastle University
janetta.steyn@ncl.ac.uk

Tom Komar
Newcastle University
tom.komar@ncl.ac.uk

Jeremy Bryans
Coventry University
jeremy.bryans@coventry.ac.uk

v1.1.1
2018/09/10

ABSTRACT

This paper discusses the development of the CEM-DIT (Communication in Emergency Management through Data Integration and Trust) system, which allows decision makers in crises to send out automated data requests to multiple heterogeneous and potentially unknown sources and interactively determine how reliable, relevant and trustworthy the responses are. We describe the underlying technology, which is based partially on data integration and matching, and partly on utilisation of provenance data. We describe our cooperation with the Urban Observatory (UO), which allows us to develop the system in collaboration with developers of the kind of crisis-relevant data which the system is designed for. The system is currently in development, and we describe which parts are fully implemented and which are currently being developed.

Keywords

Provenance, Data matching, Data integration, Semantic web systems, Decision support

INTRODUCTION

It is well understood that responses to crises are often hampered by the inability to share data quickly and efficiently. Living in a world of data means not only that huge data sources are available but also that huge numbers of different data sources exist, many of which may contain data that is relevant to any given situation. But these data sources are generally heterogeneous, may be hard to understand and derive from a wide variety of sources, some of which may be unknown and/or untrusted. The traditional approach to data exchange during crisis management - largely through direct human interaction - is simply unable to access and process these data sources in depth, which can mean that important and relevant pieces of information are unavailable to decision makers at the crucial time. In order to make full use of all this data, automated and interactive intelligent systems are required to discover, filter, interpret and annotate potentially relevant data for efficient and effective consumption by decision makers. The fact that such

*corresponding author

systems are not in place is a reflection of the difficulty of designing them. There are several technical difficulties around this problem; two pressing ones that require intelligent solutions are *(i)* How to deal with data heterogeneity? When data in different resources will use different terminology, different structure, different formats and so on, how can this data be made mutually comprehensible and how can decision makers understand how data from some source is relevant to what they want to know when it might appear very different? Secondly *(ii)* when data is derived from multiple sources, some of which are unknown, how can decision makers trust that data? How can they find out what they need to know about where the data comes from in order to be confident in making decisions based on it? It is crucial for decision makers that the data they receive is believable, can be trusted and is of an acceptable quality (Prat and Madnick 2008). Provenance provides transparency into the sources and processing history of the data (Alper et al. 2016).

In this paper, we describe the CEM-DIT system, which will allow decision makers to send out automated queries to multiple data sources, and interactively explore potential responses to their queries in order to determine whether or not the responses meet their requirements. The system is based on extensive work by the authors in data integration and in provenance, and focuses on integrating these two orthogonal issues into a unified system that priorities the autonomy of the decision makers to explore options and explanations and thereby to make decisions with confidence.

The theoretical underpinning of the system is well developed and the system already exists in part. However, full integration and design decisions remain to be done, and we present this work in progress to the ISCRAM community in the hope of getting feedback and interaction that will be helpful in guiding this process.

In the paper, we first outline the related work to show how our approach extends the state of the art. Next we describe the CEM-DIT system, and then describe the underlying technology in a little more detail. We describe our data-driven development and evaluation, and finally outline our next steps and conclude the paper.

RELATED WORK

In this section, we describe how the overall approach compares to other attempts to address the problem, and then explain how the underlying technology is particularly suitable for this task, and to describe the concept on which the system is based.

Data-driven decision making support

During crises, communication must be created and maintained between different organisations, as well as between organisations and the public. A considerable amount of humanitarian data is widely available, including open data, which can be of value to decision makers. However the data formats vary massively and single problem context requires identifying and extracting the relevant and meaningful parts from many systems for representation (Paulus et al. 2018)

Much current research focusses on communication between crisis management organisations and the public, on creating channels for communicating necessary information to the public and using information from crowdsourcing and social media to contribute to the success of emergency management (Freitas et al. 2016), (Ruiz-Zafra et al. 2014). Trust and provenance are significant issues in using such data. Interoperabilty is also recognised as a problem in communication between organisations (Fischer et al. 2016). There is a lack of suitably structured information (Rietjens et al. 2009), and when structured information exists the terminology, formats and structures vary between organisations. While there is work at an organisational level on creating co-operation between agencies, our work focuses on data interoperability between agencies at a semantic level, and on dynamic use of disparate data so that data sources not previously identified and integrated before runtime can be utilised.

Query Rewriting and Provenance

The CEM-DIT system is based on dynamic structural semantic query rewriting, as well as user-focussed provenance encapsulation and exploration. To the best of our knowledge there are no other approaches using failure-driven reasoning techniques combined with matching techniques to dynamically rewrite queries. Work by (Quesada Real et al. 2017) and (Shamoug et al. 2018) addresses the differences in the ways language is used in different organisations which make information sharing more complex. Ontologies offer an approach to linking the different terminologies within organisations and to represent explicitly both differences and similarities in the way that terms are used and understood (Galton and Worboys 2011; Burel et al. 2017). Ontologies have been used to support decision makers in crisis management in different organisations. (Burel et al. 2017) derives an ontological model to represent crises and demonstrates its use to support document sharing for crisis management. RESCUER (Simas et al. 2017) provides semantic interoperability in crisis management by applying ontology-based schema matching to integrate crowdsourced data with more detailed data from pre-existing systems in command and control. Our approach is similar to that of RESCUER, but differs in that we use structured semantic matching in addition to ontology matching, which allows us to match not only the terminology but the structure of a query, and we apply them to query response rather than data integration.

Data provenance is a structured form of metadata designed to records the activities and datasets involved in data production, as well as their dependency relationships. The PROV data model (Moreau et al. 2013), released by the W3C in 2013, defines a schema and constraints that together provide a structural and semantic foundation for provenance. This enables the interoperable exchange of provenance between data producers and consumers. In addition to containing references to data generation or transformation processes, a *provenance trace* typically includes input or intermediate data products as well as references to *agents*, that is the humans or software systems who were responsible for enacting those processes. In multi-party collaboration settings that involve data sharing, as well as in third party auditing of data and processes, there is a broad expectation that shipping the available provenance to collaborators, or more generally publishing it along with the data, may help data consumers, including auditors, form judgments regarding the reliability of the data itself.

There is however a tension between the potential for using provenance to establish data reliability, and its complexity. Provenance-aware information systems are equipped with provenance-capture capabilities, which in many cases allow fine-grained, detailed provenance traces to be collected automatically by the underlying system, cf. (Gehani and Tariq 2011). Provenance that is collected about high-level processes such as workflows is less detailed, but still not suitable to be understood by humans, especially decision makers on the receiving end of query results, who are not familiar with the systems that generated the data in the first place. For the same reason, it is not reasonable to assume that the same decision makers will be able to query the provenance in order to extract its expected value. it is therefore necessary to simplify the provenance content in order to make it understandable by these users. At Newcastle we have addressed this problem in the context of selective disclosure of provenance, i.e., when provenance that is exchanged between parties with limited mutual trust needs to be partially *obfuscated* (Missier et al. 2014). In this work we adopt a similar approach and reuse some of the same ideas, namely to create views over the raw provenance trace, this time with the aim to support interactive exploration of complex provenance graphs by decision makers.

The CEM-DIT system not only utilising this matching and provenance technology, but in addition the integration of these technologies is unique. This allows us to create a system where a decision maker can not only access heterogeneous data from multiple disparate data sources, but can be provided with the tools to appropriate evaluate the data, and hence make well-justified decisions, addressing two crucial and orthogonal issues in data sharing.

AN INTEGRATED SYSTEM FOR DECISION MAKERS

In this section, we outline what the CEM-DIT system will ultimately look like, highlighting what is already in place and what remains to be done.

We illustrate CEM-DIT through a scenario in which there has been a major unexpected flood in the city centre. Emergency services want to enter the central area by the fastest route, while people and traffic are leaving by many routes. The emergency services need to know who is still in the area and to plan routes which are open and not obstructed by the flood or by people leaving. Traffic levels in some parts of the central area are routinely monitored by local traffic police using CCTV and electromechanical sensors. In addition, a new service is available which monitors traffic in some locations using satellite imagery. However this service is under development and is not fully integrated. These different sources of traffic data are offered in different locations by services using different formats and interfaces. The CCTV and satellite image data are automatically interpreted using image analysis software and complex machine learning algorithms. Data from different sources is not always available and new algorithms may not produce fully reliable data. The CEM-DIT system gathers, processes and returns potentially relevant data to the decision maker, together with the metadata necessary for the decision maker to make an informed decision. Figure 1 illustrates the workflow this involves.

STEP 1: Data is requested

The decision makers send out an automated query[1], written in a formal querying language. The query can either be sent out to a list of known data providers, or can be broadcast to anyone interested in responding.

In our running example, we consider a query in which the decision maker needs to know the traffic levels which have been recorded by each of the sensors in an area at a given time, with their identifiers and locations.

[1]This can be pre-written prior to interaction for predictable queries. For non-predictable queries, we are investigating systems that automatically generate formal queries from natural language text. In the meantime, non-predictable queries will require the presence of someone (the decision maker or someone else) who is able to create queries.

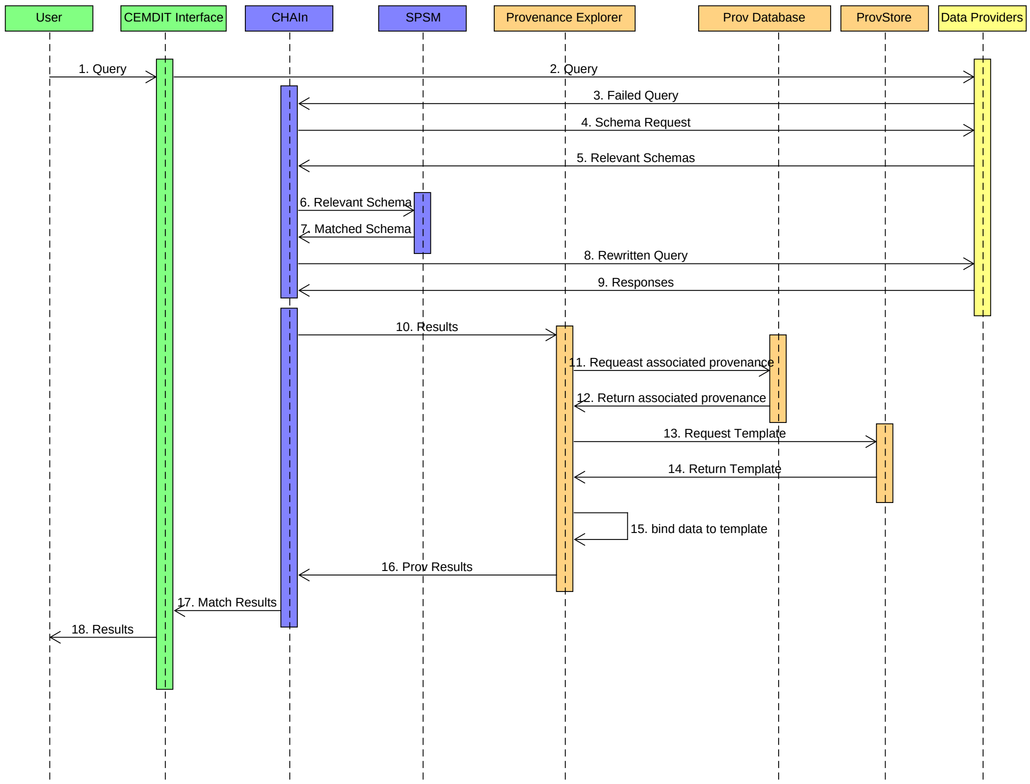

Figure 1. Sequence diagram showing the flow of data in the CEM-DIT system

STEP 2: CEM-DIT retrieves and sorts relevant data

- Multiple data sources receive the query. The query will be written according to a certain schema and will use certain terminology . Unless extensive pre-alignment has taken place, it is unlikely that the receiving data source will use precisely the same schema and/or terminology, and therefore the query will fail, even if there is data in the data source that is relevant. This is illustrated by the example query which initially involves six arguments in a given order, and talks about *latitude*, whereas relevant data sources may use the term *sensor centroid latitude* for a similar or identical concept, and may have additional or missing attributes. At this point, the CHAIn (Combining Heterogeneous Agencies' Information) matching system (McNeill et al. 2014) is initiated. It filters the data source for anything potentially relevant, and then perform detailed structural-semantic matching on potential candidates. This matching provides a map of how to rewrite the query in such a way that it is similar to the original query but capable of returning data from the data source. The process determines: 1) how similar, indicated by a score $S \in [0..1]$, the rewritten query is to the original query; b) what assumptions, rewritings and removals have been used to create the rewritten query. Any rewriting with a score S that passes a given threshold S_t is sent to the data source and the results returned to the CEM-DIT system.

- When the data is returned in response to the rewritten queries, it is annotated with any provenance data associated with it. This may be simple metadata such as the data source it came from and when it was generated, but may be much more complex - for example, complete pipelines which include applications and parameters that were used to produce the data. ProvExplorer is the part of the CEM-DIT system that extracts and visualises provenance.

  ProvExplorer requests access to the database that associates a query with locations where provenance is stored. The database also contains the ID of a provenance template which is stored in ProvStore[2].

STEP 3: Data is presented to the decision maker and evaluated

[2]ProvStore is a web service for storage of provenance documents (`https://openProvenance.org/store`)

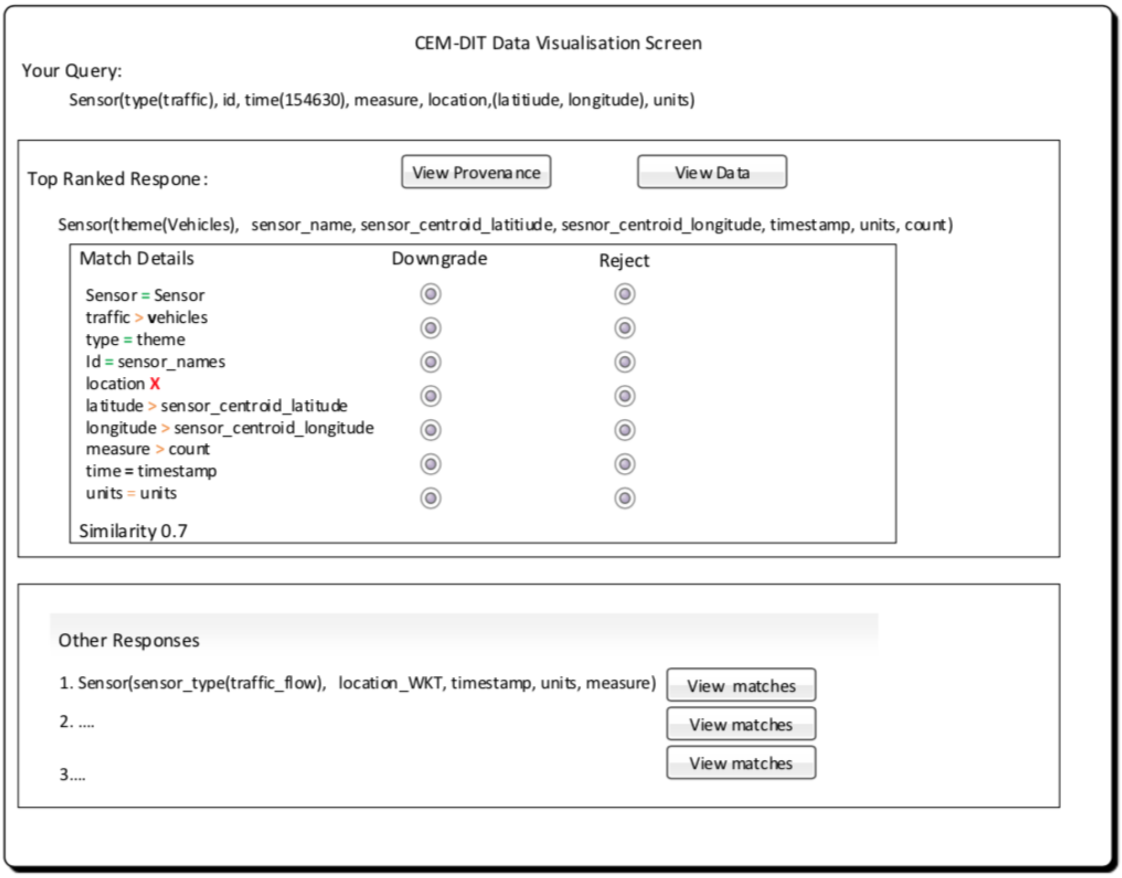

Figure 2. Mock up of a response screen for a query

The annotated responses are presented to the decision makers through the CEM-DIT system in a way that allows a decision maker to visually explore the potential responses. This step is currently under development. Here, we outline what the system is going to look like.

The CEM-DIT system presents the different options for query rewriting (if more than one passed the threshold) to the decision maker, clearly indicating where substitutions, approximations and non-matches have been made. Figure 2 shows a mock-up of the screen the decision maker would see after sending out the example query asking for data from sensors that are monitoring traffic at a given time stamp 154630. Although the query will have been sent to the data source in SQL or a similar querying language, it is presented to the decision maker as a first-order term, stripping out the querying language syntax, as this is easier for the decision maker to understand. The example query described in Step 1 will be rendered as follows:

sensor(type(traffic), id, time(154630), measure, location(latitude,longitude), units).

The schema of the top ranked response is shown, and a list of the matching of every aspect of the original query is given. The structural and semantic similarity between the top-scoring query and data source is reflected in the relatively high match similarity at 0.7. The score is lowered by differences in terminology and structure. For example, *traffic* and *measure* are more general terms than *vehicles* and *count*; a nested structure for *location* in the query is missing from the data source although its components (*latitude* and *longitude*) are both identified. Any components that are present in the data source but are not used in the query do not affect the match score. The decision maker can choose to downgrade or reject any of these matches, and if this is done it will cause all matches to be reranked, with all containing the unacceptable match either downgraded or removed from the list (depending on the type of downgrading chosen). Downgrading a match means penalising a particular aspect of it more heavily. For example, in the match details in Figure 2, we can see that a match *ID = sensor_names* is considered to be equivalent by the matching system, which means the match from the one to the other does not cost anything (i.e., reduce the overall similarity score), but the decision maker may consider that in this context they are related but not identical - hence choosing the 'downgrading' option - or that they are not related - hence choosing the 'reject' option. The former choice will add a cost to removing the match, and the latter will rule it out as a possibility, leading to a different (reduced) matching score. If this match also occurs in other schema matches, then similar alterations to the

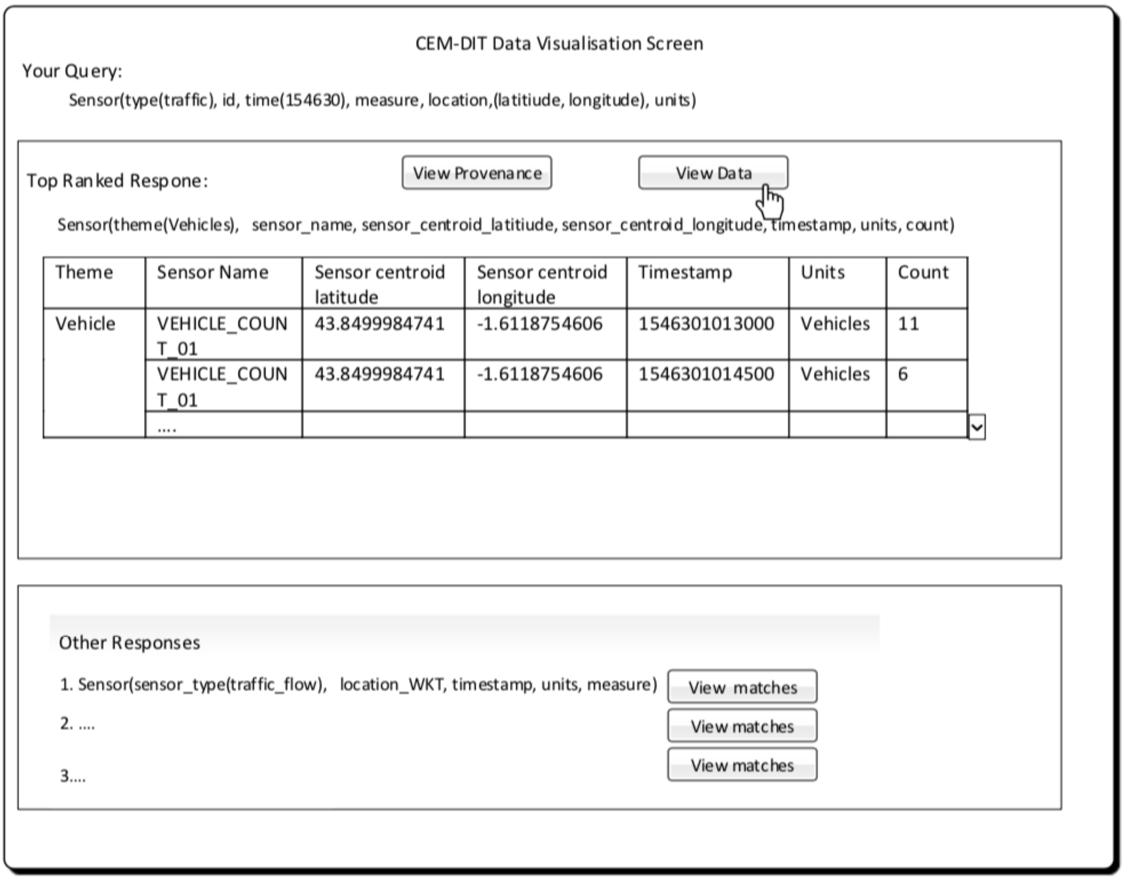

Figure 3. Mock up of data associated with the top-ranked response

cost will also occur in those matches, potentially leading to a reranking of the returned results. The decision maker can see the data associated with the top-ranked response by clicking on the *view data* button, as illustrated in Figure 3. The decision maker can also choose to view more detail about other matches lower down the ranking.

The decision maker can choose to view and explore the provenance associated with each aspect of the top schema. Initially a simplified provenance graph is presented, as illustrated in Figure 4. We can see some of the provenance of *count*, which is based on a model and an image. In order to feel confident of the data, the decision maker may then choose nodes in the provenance graph to explore further, to get more insight into how *count* is generated. Figure 5 illustrates the graph the decision maker would see if she expanded both *model* and *count* (this is shown outwith the CEM-DIT screen to improve readability). Any node highlighted in red can be expanded to expose more provenance data; for nodes that are not highlighted there is no more additional provenance data available. This additional provenance may be lacking either because data owners have not have recorded or stored full provenance for the data, or because they are unwilling to share all aspects of provenance for commercial or privacy reasons. If provenance data that the decision maker wants to see is not available, she can downgrade the match on that basis. As with the matching downgrades, in the finished system this information will be propagated to other matches, and any that do not contain essential provenance information will be removed.

Once the decision maker is confident that a particular result is acceptable from both a matching and a provenance point of view, she is able to accept it and the data attached to that result can be used, either automatically, interactively or by hand, in any decisions to be made.

This step is partially implemented. The key steps that are still to be implemented are outlined in the Next Steps section.

UNDERLYING TECHNOLOGIES

The core functionality of the CEM-DIT system depends on our work on provenance and on query repair. Here, we briefly explain these technologies.

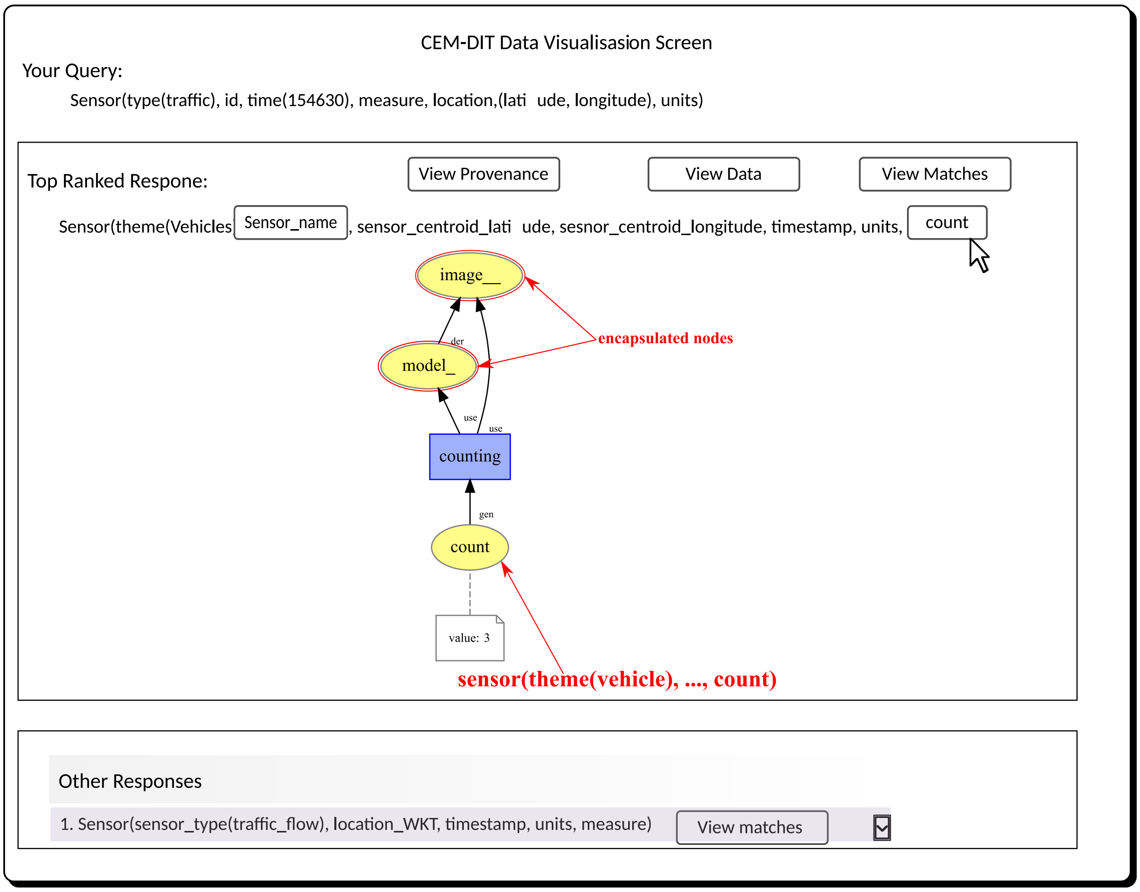

Figure 4. A simplified provenance model which is shown to the decision maker in the first instance. The original query is shown in red with an arrow pointing to the result for count.

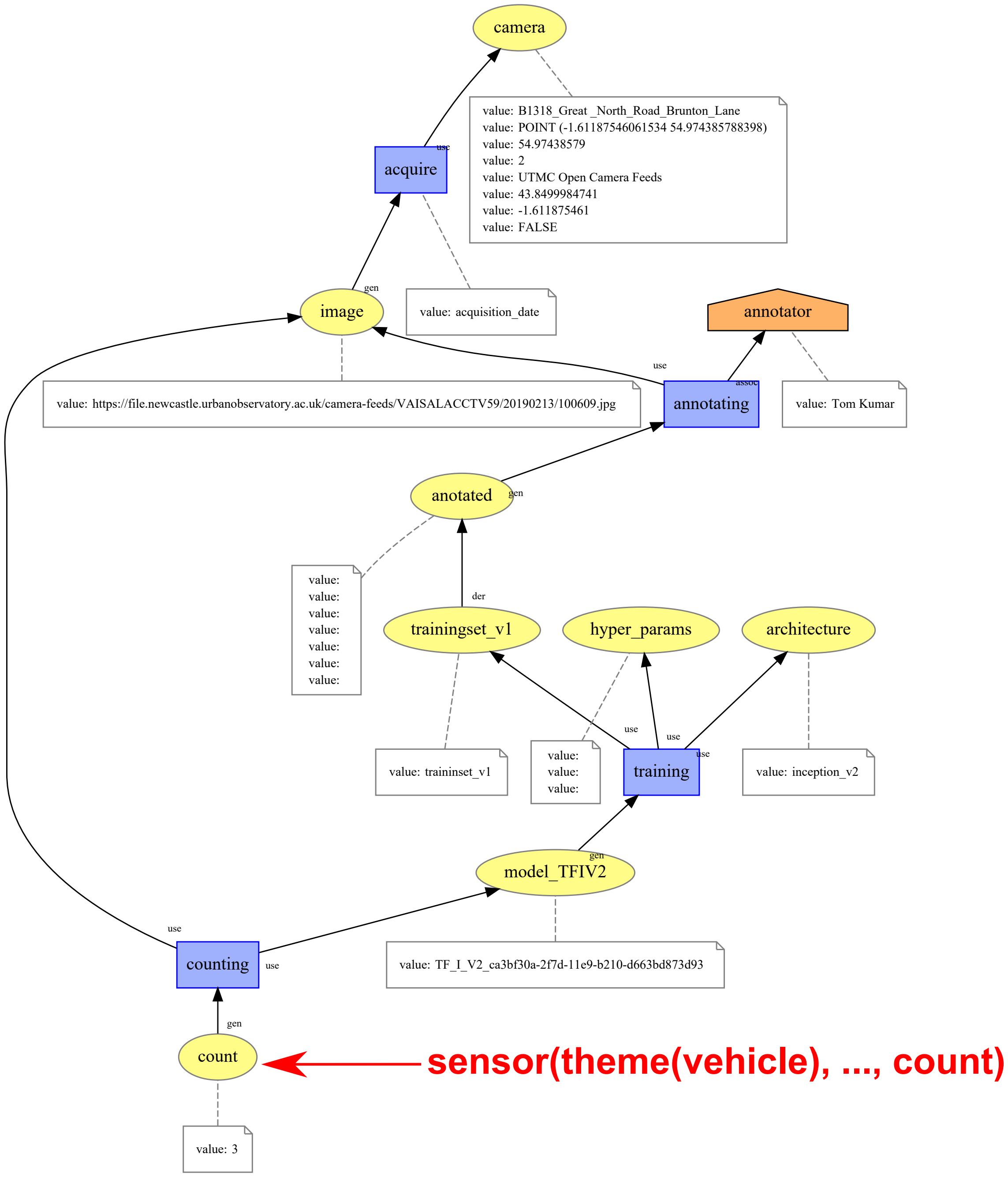

Figure 5. The provenance graph of all provenance retrieved for the results. Where 'value:' is shown, in the attributes, with no value, the provenance was not available.

ProvExplorer

Provenance can be provided as meta-data when queries are made. When provenance is not available directly with the data it may also be obtained from a separate provenance store which is maintained by the data provider or a third party such as ProvStore. By mapping the retrieved provenance to entities, activities and agents, according to the W3C provenance standards (https://www.w3.org/standards/techs/provenance#w3c_all), we are able to visualise it as a provenance graph. Such graphs can then be manipulated to encapsulate the provenance. To simplify and automate the process of adding and retrieving provenance for data obtained from queries we have developed an application, ProvExplorer, to serve as a workbench for the retrieval and exploration of provenance. The application is written in Java and makes use of the ProvToolbox API (Moreau 2019). ProvExplorer offers a GUI with a PROV-N (a human readable provenance notation) editor. The PROV-N can be viewed as a provenance graph. To encapsulate provenance data, operators can then be applied as policies.

The current system facilitates provenance users encapsulating (and thereby simplifying) their own provenance data. The potential to obfuscate (hide) provenance data will still be an important part of the process on the data owner's side, but the central focus of the system - that of providing relevant data to decision makers - requires the automation of graph simplification and the facility to expand simplified nodes as desired.

Query Repair

When a query fails, CHAIn uses structural semantic matching to repair the query with respect to the target data source and the queries are rewritten accordingly (McNeill et al. 2014; Giunchiglia et al. 2008). The matching part of the process is done by the SPSM algorithm (Giunchiglia et al. 2008), which performs pairwise matching of structured terms. SPSM combines structural matching with ontology-based matching, where existing ontologies and lexicons, such as SUMO (Niles and Pease 2001) and WordNet (Fellbaum 1998), as well as domain-specific ontologies where available, are used to determine how similar different words and terms are. Matching can usually be done in multiple ways and matches will rarely be perfect, so the similarity score is calculated to indicate the closeness of each match. For each match SPSM returns the similarity and a set of mappings between the different components of the two terms. The mappings are used to write a repaired version of the query which can be run on the datasource, and the repaired query, similarity score and mappings are returned.

SPSM is an expensive process, and it is not feasible to do it between a single query and every possible datasource. CHAin therefore performs a filtering step first, narrowing down the datasource schemas to a subset of likely candidates. Related terms are computed based on the SUMO and WordNet ontologies. In Figure 2 the schemas for the query and for the top ranked responses are all represented by the same predicate "sensor", and therefore these data schemas pass the narrow-down step and proceed to SPSM matching.

The schema-matching part of the process matches first-order terms: that is, terms of the form
$predicate(Arg, Arg, ..., Arg)$,
where each Arg may itself be a function. This process is therefore applicable to data formats that can be translated into this format. This is a very general format which includes a vast number of common representations. Currently, the CHAIn system can be used for SPARQL queries to RDF datasources and for SQL queries to RDB datasources, and may be extended to hierarchical and tabular structures such as JSON and spreadsheet tables.

DATA-DRIVEN DEVELOPMENT AND EVALUATION

It is important to develop the system with reference to plausible - ideally real - data that gives scope for both the matching and provenance side of the system. This allows us to base the development on plausible case studies and to carry out realistic evaluations on the resulting system. Currently, there is a paucity of data sources which contain deep, interesting provenance data. In part this is because tools to exploit that provenance are only now being developed. Fortunately, we have been able to work with the Newcastle Urban Observatory[3] (UO). The UO is an ongoing UKRIC project involving, among others, Newcastle University that provides the largest set of publicly available real time urban data in the UK. The Newcastle project covers city centre and the suburbs. More than 50 data types organised in 12 themes such as air quality, traffic, weather, are gathered from sensors across the city. A RESTful API is made available for developers allowing data to be downloaded in JSON or CSV format.

Although the UO does not address the issue of capturing provenance specifically, it does capture some data that can be used to verify the quality of data. As per our running example, we are working primarily with sensor data for sensors that are involved in counting cars and people. The process used for producing the counts make for an interesting pipeline for which the quality of data might be questioned by decision makers. For instance, a query can

[3] http://www.urbanobservatory.ac.uk/

be made for the number of cars that were detected by a sensor in a specific location at a specific time. Questions such as the following might come to mind when the results are retrieved:

1. How are these counts generated?
2. How are the counts affected by adverse weather and light conditions?
3. What happens if one car partially obscures another?

As well data as with good provenance data attached, we need data that comes from multiple similar data sources so that we can develop and evaluate the matching aspects of the system. The UO supports three different opportunities for data matching. First, there is some variability within the UO data sources. The UO gathers data from different sources. Sensors are of different types and similar data is returned from sensors provided by different services with varying data structures and semantics. In our use case information about traffic flow may be obtained from cameras, satellite images or electromechanical sensors, in different but possibly overlapping locations. Second, the generality of the CEM-DIT system can be tested by posing queries which are based on other city data sources such as NYC Open Data Repository[4], and attempting to repair these queries with CEM-DIT so that they run on UO data. Third, evaluation can be performed via synthetic queries which mimic the variability of real-life queries. We have developed a system, CRPTR for CHAIn, which takes queries which will run on an existing knowledge source and generates wide variety of plausibly corrupted queries. For evaluation we will create multiple corrupted queries and attempt to repair them with CEM-DIT so that they run on UO data.

Evaluation Aims

Evaluation for a system such as CEM-DIT is a tricky business. Ultimately, we want to ask the question *when decision makers are able to access a wide variety of data through the CEM-DIT system, and use the system to facilitate their understanding of and trust in this data, do they make better decisions than if they do not have access to such automated support tools?* Quite apart from determining how we evaluate 'better' decisions, direct access to decision makers in a simulated context that is meaningfully similar to a real crisis situation is expensive and difficult. It is essential that we carry out an evaluation that demonstrates the usability and value of the system before it is used in simulations with real responders.

Our concern at the moment is to evaluate two aspects of the system: (*i*) does the underlying technology provide results that are appropriate; (*ii*) does the system allow decision makers to interact with these results in a useful way.

- (*i*) Determining whether the results returned are appropriate means:
  - *are the matches returned relevant to the queries asked, and is the ranking a good reflection of the differing quality of the match?* Determining this is not straightforward as, in most cases, there is no gold standard against which we can judge our results, and there is some amount of subjectivity in assessing whether one rephrasing of a query is 'better' than another. There are various attempts in the matching community to create gold standards for evaluation, for example the Ontology Alignment Evaluation Initiative[5]. Whilst this initiative does not address the same matching problem as we are, we can use the test sets generated by this initiative to support evaluation. In addition, CRPTR for CHAIn corrupts data sets in a variety of ways to create plausible mismatched sets with a pre-existing gold standard, and hence facilitates evaluation.
  - *is the decision maker able to immediately access a simple version of the provenance data and then, if desired, use the system to access all the provenance information available?* Evaluating this is a much simpler prospect, requiring a definition of what an acceptable initial compression of the provenance data would be, and a demonstration that access to complete available provenance data was always possible.
- (*ii*) Since developing a sophisticated GUI that would be usable in the field is outside the scope of our current work, investigating the second point will involve getting decision makers who understand the data, and in particular the provenance, to give us feedback on the usability of the system.

We are currently working with the UO to develop a suitable evaluation scenario with multiple mismatched data sources and relevant accompanying provenance data generated.

[4] https://opendata.cityofnewyork.us/
[5] http://oaei.ontologymatching.org

NEXT STEPS

The majority of the underlying technology for the system described above is already implemented, but some steps remain of both theoretical development and implementation before the system described above is fully realised.

- **Extending the GUI to support the full functionality of the system.** The current project will include extending our existing GUI, which focusses on interaction with provenance data, to allow full visualisation of the matching data as illustrated in Figures 2 and 3. Creating a GUI suitable for use in the field is outside the scope of our current project but working closely with decision makers to understand the kind of interface they would need to use such a system is an essential part of bringing such technology to the field and something we intend to focus on in future.
- **Extension of the matching system to facilitate the dynamic reorganisation of matches based on user feedback.** The scoring algorithm of the matching system needs to be updated so that when matches are downgraded by the decision maker the downgrade is propagated to other relevant matches, and so that potential matches are removed automatically from the returned results list if they include aspects that decision makers have indicated are unacceptable.
- **Inverting the encapsulating functionality in the provenance system.** ProvExplorer has been designed as a tool for those who own provenance data to be able to control it, allowing them to encapsulate any parts they do not wish to view in detail. CEM-DIT requires this functionality to be inverted, as it requires a simple version of the provenance data to be presented to the user, with the ability to expand that as far as the user requires, or is possible.

CONCLUSIONS

In this paper, we introduced the CEM-DIT system, which is designed to support decision makers during crises by allowing them to easily access, understand and trust data from multiple heterogeneous data sources, which may be unknown and/or partially trusted. The theoretical underpinnings of the system and significant components of the implementation are in place, and we are currently working on the full implementation of the system.

REFERENCES

Alper, P., Belhajjame, K., and Goble, C. A. (2016). "Automatic vs manual provenance abstractions: mind the gap". In: *arXiv preprint arXiv:1605.06669*.

Burel, G., Piccolo, L., Meesters, K., and Alani, H. (2017). "[D]o[RES] - A Three-Tier Ontology for Modelling Crises in the Digital Age". In: vol. 2017-. Information Systems for Crisis Response and Management, ISCRAM, pp. 834–845.

Fellbaum, C., ed. (1998). *WordNet: an electronic lexical database*. MIT Press.

Fischer, D., Posegga, O., and Fischbach, K. (2016). "Communication Barriers in Crisis Management: A Literature Review". In: *24th European Conference on Information Systems, ECIS 2016*. Association for Information Systems.

Freitas, D. P., Borges, M. R., Vivacqua, A. S., and Carvalho, P. V. (2016). "Communication Channels Among Various Roles During Crisis Response". In: Information Systems for Crisis Response and Management, ISCRAM.

Galton, A. and Worboys, M. (2011). "An ontology of information for emergency management". In: *Proceedings of 8th International Conference on Information Systems for Crisis Response and Management*.

Gehani, A. and Tariq, D. (2011). "SPADE: Support for Provenance Auditing in Distributed Environments". In: *SRI International*.

Giunchiglia, F., McNeill, F., Yatskevich, M., Pane, J., Besana, P., and Shvaiko, P. (2008). "Approximate Structure-Preserving Semantic Matching". In: *On the Move to Meaningful Internet Systems: OTM 2008*. Springer, pp. 1217–1234.

McNeill, F., Gkaniatsou, A., and Bundy, A. (2014). "Dynamic Data Sharing from Large Data Sources". In: *Proceedings of 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2014)*.

Missier, P., Bryans, J., Gamble, C., Curcin, V., and Danger, R. (2014). "ProvAbs: model, policy, and tooling for abstracting PROV graphs". In: *International Provenance and Annotation Workshop*. Springer, pp. 3–15.

Moreau, L. (2019). *The Open Provenance Website*.

Moreau, L., Missier, P., Belhajjame, K., B'Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., et al. (2013). *PROV-DM: The PROV Data Model*. Tech. rep. World Wide Web Consortium.

Niles, I. and Pease, A. (2001). "Towards a Standard Upper Ontology". In: *Proceedings of the International Conference on Formal Ontology in Information Systems - Volume 2001*. FOIS '01. Ogunquit, Maine, USA: ACM, pp. 2–9.

Paulus, D., Meesters, K., Walle, B. van de, Boersma, K., and Tomaszewski, B. (2018). "Turning Data into Action: Supporting Humanitarian Field Workers with Open Data". In: *Proceedings of the 15th International ISCRAM Conference*.

Prat, N. and Madnick, S. (2008). "Measuring Data Believability: A Provenance Approach". In: *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, pp. 393–393.

Quesada Real, F., McNeill, F., Bella, G., and Bundy, A. (2017). "Improving Dynamic Information Exchange in Emergency Response Scenarios". In: *Proceedings of 14th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2017)*.

Rietjens, S. J., Verlaan, K., Zaalberg, T. W. B., and De Boer, S. J. (2009). "Inter-Organisational Communication in Civil–Military Cooperation During Complex Emergencies: A Case Study in Afghanistan1". In: *Disasters* 33.3, pp. 412–435. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-7717.2008.01081.x.

Ruiz-Zafra, A., Núñez, A.-G., Penadés, M., Canós, J., and Borges, M. (2014). "[SUCR]E: Supporting Users, Controllers and Responders in Emergencies". In: *ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management*. The Pennsylvania State University, pp. 255–259.

Shamoug, A., Cranefield, S., and Dick, G. (2018). "Information Retrieval for Humanitarian Crises via a Semantically Classified Word Embedding". In: vol. 2018-. Information Systems for Crisis Response and Management, ISCRAM, pp. 132–144.

Simas, F., Barros, R., Salvador, L., Weber, M., and Amorim, S. (2017). "A Data Exchange Tool Based on Ontology for Emergency Response Systems". In: *Research Conference on Metadata and Semantics Research*. Springer, pp. 74–79.