

**Impact case study (REF3b)**

<p><b>Institution:</b> University of Sheffield</p>
<p><b>Unit of Assessment:</b> 11 – Computer Science and Informatics</p>
<p><b>Title of case study:</b> OAK: Harnessing the power of information for situation awareness and organisational intelligence</p>
<p><b>1. Summary of the impact</b></p> <p>Researchers in the Organisations, Information and Knowledge (OAK) group have developed technologies for large-scale acquisition, integration and sense-making of information acquired from a variety of sources, including textual documents, the Web and multiple devices. These technologies have had:</p> <ol style="list-style-type: none"> <li>1. <i>Economic impact</i> in form of two University spin-out companies, created in order to exploit them: <b>K-Now Ltd</b>, who use the technologies to support knowledge management in large enterprises and social media monitoring for emergency response, and <b>The Flow Ltd</b>, who use the technologies to power organisational intelligence in, e.g. telematics-based motor insurance.</li> <li>2. <i>Economic impact</i> in the large enterprises and their supply chain that have adopted them. [Text removed for publication] have adopted the technologies as the core component of a knowledge management programme focusing on data mining thousands of documents that has saved millions and been delivered to thousands of engineers; Direct Line are offering driver-behaviour-based motor insurance to [text removed for publication] customers based on the technologies.</li> <li>3. <i>Public service impact</i> by using them for social media monitoring to deliver improved civil monitoring and protection services for hundreds of thousands of people, e.g. at large public festivals [text removed for publication], and for river flood monitoring.</li> </ol>
<p><b>2. Underpinning research</b></p> <p>The OAK research group focuses on large-scale information management including:</p> <ul style="list-style-type: none"> <li>• <b>acquisition:</b> how to capture information over a large scale using multiple digital devices and in multiple modalities; examples include how to mine information from millions of documents from the Web and large distributed archives, or from thousands of social media messages a second, or from thousands of sensor inputs per second;</li> <li>• <b>integration:</b> how to make sense of the captured information: large-scale integration across archives and sources, and integration of information and its context;</li> <li>• <b>searching and sense-making:</b> how to make sense and use of the information; how to power organisational intelligence; how to present information to users.</li> </ul> <p>Since 2005, the group has received funds of around £6.5M from the EPSRC, the EU, AHRC, MRC, JRC and TSB; about 10% came directly from industry.</p> <p><b>Acquisition:</b> The underpinning research started in 2000 as part of the EPSRC IRC Advanced Knowledge technologies (AKT) project. This project was organised around the concept of integrated-support knowledge management (KM) in organisations, covering the whole knowledge lifecycle, from information capture, to integration, visualisation and sense-making. Starting in 2000, <i>Prof Ciravegna</i> (Sheffield since 2000) carried out pioneering research into how machine learning could be applied to the problem of mining information from natural language texts. This research resulted in novel techniques for user-centred, machine learning-based information extraction [R1]. From 2007, <i>Ciravegna</i> and team extended these techniques and applied them to mining the Web and Social Streams (e.g. Twitter and Facebook - EU projects WeKnowIt and WeSenseIt [R3]). These studies supplied insights that enabled the development of new techniques for automated acquisition of knowledge from a large number of distributed sources. This in turn allowed applications of KM in large enterprises and for monitoring the social web. Knowledge acquisition from mobiles was studied by <i>Lanfranchi, Chapman</i> and <i>Ciravegna</i> in 2007-2012 within the European Project WeKnowIt. This formed the basis of the acquisition of behavioural information, which led to the creation of <b>The Flow</b>, and the large-scale monitoring of social media for</p>

## Impact case study (REF3b)

emergency response and situation analysis.

**Integration:** Starting in 2004, the OAK group worked on methods for large-scale integration of information: among them the approaches to integration of large-scale Web resources summarized in [R5]. In 2005 they created the most widely used library of string metrics in the world (SimMetrics – 65,000 downloads since 2005 [S6]). String metrics are methods to efficiently and effectively match equivalent database records and textual descriptions over a large scale. They were the foundations that led to the development of methods for Terminology Recognition (TR) in technical domains currently in use at Rolls-Royce for KM purposes. By building on SimMetrics and further research on Information Extraction [R4], TR allows information from different sources to be searched for and located using a single request, rather than requiring separate searches of multiple systems. Moreover, since 2009, *Ciravegna* and his RAs have worked on integration of information extracted from text, images and GPS enabled devices with linked open data. This is the underpinning technology commercialised by **The Flow**.

**Searching and Sense-Making:** In 2005-2010, *Baghdev, Chapman, Lanfranchi* and *Ciravegna* created the Hybrid Approach paradigm to searching distributed archives [R6]. In 2007-2009, *Ciravegna* et al. closed the information lifecycle by studying ways of visualising and making sense of large scale distributed information for organisational purposes. Sense-making was based on user-centred multi-visualisation of data with dynamic filters [R2]. This technology is currently used in the visualisation and sense-making of millions of tweets in emergency response applications and as a foundation technology by their spin-out company **K-Now**.

### 3. References to the research (\*\*\*) denotes outputs which best demonstrate underpinning research quality)

- R1.** F. Ciravegna. *Adaptive Information Extraction from Text by Rule Induction and Generalisation*. In Proceedings of 17th International Joint Conference on Artificial Intelligence (IJCAI 2001), Seattle, August 2001.
- R2.** \*\*\* D. Petrelli, S. Mazumdar, A.-S. Dadzie and F. Ciravegna, *Multi Visualisation and Dynamic Query for Effective Exploration of Semantic Data*, In Proceedings of the 8th International Semantic Web Conference, Chantilly, Virginia, October 2009. This paper received an “Honourable Mention” award at the conference. doi: [10.1007/978-3-642-04930-9\\_32](https://doi.org/10.1007/978-3-642-04930-9_32)
- R3.** \*\*\* M. Rowe and F. Ciravegna, *Disambiguating Identity Web References using Web 2.0 Data and Semantics*, in International Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 8 (2), pp. 125-142, 2010. doi: [10.1016/j.websem.2010.04.005](https://doi.org/10.1016/j.websem.2010.04.005)
- R4.** J. Iria, N. Ireson and F. Ciravegna, *An Experimental Study on Boundary Classification Algorithms for Information Extraction using SVM*, in Proceeding of the 11th Conference of the European Chapter of the Association for Computational Linguistics, April 2006.
- R5.** Z. Zhang, A. Gentile and F. Ciravegna: *Harnessing different knowledge sources to measure semantic relatedness under a uniform model*. In Proceedings of the International Conference on Empirical Methods in Natural Language Processing (EMNLP2011), Edinburgh, July 2011.
- R6.** \*\*\* R. Bhagdev, S. Chapman, F. Ciravegna, V. Lanfranchi and D. Petrelli, *Hybrid Search: Effectively Combining Keywords and Semantic Searches* in Proceedings of the 5th European Semantic Web Conference, ESWC 08, Tenerife, June 2008. doi: [10.1007/978-3-540-68234-9\\_41](https://doi.org/10.1007/978-3-540-68234-9_41)

### 4. Details of the impact

The technologies developed at Sheffield have enabled a variety of impacts of which three principal ones are: (i) *economic impact* of KM technologies in large distributed organisations (Terminology recognition currently in use at [text removed for publication] and Sheffield spin-out **K-Now**); (ii) *economic impact* via monitoring of driver behaviour for motor insurance pricing using mobile phones (commercialised by Sheffield spin-out **The Flow**); and (iii) *public services benefit* via the monitoring of social media for emergency response (large public events/flood monitoring).

#### Knowledge Management in Multinational Organisations

Our technology for acquisition, integration and sense-making for KM within large organisations, as

## Impact case study (REF3b)

described above, has led to economic impact via two routes. First, the technologies have been applied and refined for use within [text removed for publication], leading to substantial economic impact within that organisation. Our Terminology Recognition model, algorithm and software [R4] were developed as part of a [text removed for publication] CASE studentship supervised by *Ciravegna*. They were further developed by [text removed for publication], who subsequently hired the student. The technology was certified for use within [text removed for publication] in February 2012. TR is the core component of a knowledge management improvements programme focusing on information extraction from, and data mining of thousands of documents. This enables a one-point access (e.g. via searching and visualisation) for information that would otherwise be lost in the myriad of repositories and documents. [text removed for publication] currently estimates the programme brings **cost savings of millions and has been delivered to over** [text removed for publication] **engineers** [S1]. It enables product improvement through automatic quantification of customer impact of manufacturing non-conformance. The company shortlisted TR in 2009 for the prestigious [text removed for publication] award for solutions which can sensibly change the future way of working of the company.

Second, with the strong encouragement of [text removed for publication] who wanted us to advance our approaches for large scale knowledge management beyond technology readiness levels 4-6 (the generally accepted limit for academic technology), we created a spin-out company, **K-Now**, in 2008. K-Now has commercialised part of our technology for acquisition, integration and sense-making (as described in [R6]). It now has a team of 6 software engineers and an annual turnover of £250,000. It maintains and extends the KM software for [text removed for publication] and has numerous other major customers, including KPMG, Deloitte, Adelie Foods, Comet and Associated British Food [S2].

### Driver Behaviour Monitoring

The Floop Ltd is a company spun out from K-Now Ltd (who provided the technology, CTO and Head of development) and the University of Sheffield (who provided scientific support and lead via *Ciravegna*) to commercialise telematics solutions that capture information about driving behaviour via mobile phones. Currently their turnover is over £1M and they employ 13 full-time staff in Sheffield [S3]. Their technology enables insurers to create a very detailed profile of insured drivers and hence to offer premiums that are tailored to their actual personal risk. It does this based on graph data models that utilise information from GPS sensors on an in-car mobile phone in the context of huge amounts of geographic, social and insurance data. Specifically, the technology calculates the risk of a driver's behaviour at a specific GPS-signalled location, by comparing the behaviour with all the facts known about this place (number of past incidents, traffic jams, topology, potholes, zebra crossing, schools, etc.) and the way all other drivers have behaved there (e.g. if the driver is at the speed limit, say, 30 mph but everyone else reduces to 20 mph there, then the driver's behaviour is dangerous). The technology is based on relatedness and information integration models directly derived from *Ciravegna's* group's work.

The Floop was founded in 2012 and has two major customers: Direct Line (the largest UK motor insurer) who is offering the solution to two million of their customers and AIG (Chartis), one of the largest American insurers, who is trialling the system with 30,000 drivers in Argentina, India Singapore and Israel. They have plans to extend the application to their world-wide operations [S3]. The Floop's technology provides a quantifiable reduction in risk on the side of the insurers and can dramatically reduce the premium paid by careful drivers (e.g. on average, careful young drivers reduce their premium from £2,500 to £850). Drivers with telematics car insurance policies have claims typically 30% smaller (study by Co-operative Insurance). By motivating drivers to drive more safely in order to benefit from these lower premiums, the technology is indirectly also having a positive health impact.

### Real Time Intelligence for Emergency Services via Social Media Monitoring

The real-time evolution of major events (e.g. floods, fires, protests etc.) is now being widely documented through social media. We developed a technology (TRIDS) able to monitor social

## Impact case study (REF3b)

media (Facebook, Twitter, etc.) that is being used by emergency responders, security companies, festival organisers and local councils to plan deployment of resources and respond to evolving situations. The technology automatically sifts through millions of Twitter messages per day and identifies those that are relevant to the event under consideration. It facilitates the identification of critical situations by: (a) providing access to the relevant messages; (b) visualising the contained information to give an overview picture (through trends and topics); and (c) organising messages and information according to location and timeline, as well as authors, keywords and topics.

**Impact on Public Services** A major impact of the technology is improved situation awareness before and during large public events involving hundreds of thousands of people. The new TRIDS technology, developed in the OAK group and commercialised/ marketed by K-Now, has been adopted by several public service organisations, as well as private festival organisers and is enabling improved delivery of civil/crowd monitoring and protection services. For example:

- **Bristol City Council’s Civil Protection Unit (CPU)** evaluated the TRIDS technology during the 2011 St Paul’s Festival and, following the positive outcome of that trial, invited the OAK group to support them with the technology for the St Paul’s Festival (90,000+ visitors) and the Bristol Harbour Festival (250,000+ visitors) in 2013. The Head of the Bristol CPU says: “*The use of the OAK Group technology already positively changed our vision and practices on emergency management*” [S4]. Bristol CC is now convinced that the monitoring of social media is the most effective future technology and they are heavily investing in further development together with us and K-Now via the European funded Project Eppics.
- TRIDS was used by the **Manchester Police** during the protest surrounding the Conservative Conference in October 2011, where the technology enabled monitoring social media during the event and helped inform operations in some critical situations, including the breaking away of some groups. It was also used by the **Met Police** to derive the requirements for the social media monitoring platform for the Olympics during a bomb drill at the disused underground station in Aldwych (London) in February 2012.
- The [text removed for publication] organisers used the TRIDS technology to monitor social media relating to the 2013 Festival. Their Commercial Director says: “*Information identified by the system was used by the Event Control Room to manage potential and actual events during the Festival ... (we) believe this type of technology and monitoring will become key to the management of future Festivals*”. They plan to use TRIDS to monitor the Festival for at least the next 3 years. TRIDS was also used by the organisers of the Leeds Festival in 2013.
- Doncaster Council is using the technology to help establish “Citizen Observatories” for river flooding to deliver “*more efficient, empowered and informed risk planning policy and response arrangements*”. They say: “*the technical and professional support from Sheffield University ... is having a very beneficial impact on our resilience and emergency planning policies and strategies, and is being used by emergency responders in Doncaster to harness citizen risk information and flood risk information to shape and inform our emergency response activities in a more cost effective and safe way*” [S5].

##### 5. Sources to corroborate the impact

- S1.** Letter from [text removed for publication] confirms the impact of the Terminology Recognition software within [text removed for publication].
- S2.** Letter from the Director and CTO K-Now Ltd confirms the impact of ideas stemming from Ciravegna’s research group on K-Now’s technology.
- S3.** Letter from the CEO of The Floow Ltd confirms the impact of ideas stemming from Ciravegna’s research group on The Floow’s technology.
- S4.** Letter the Head of Civil Protection Unit, Bristol City Council, confirms the impact of the TRIDS technology on their emergency management practices.
- S5.** Letter from the Resilience and Emergency Planning Officer, Doncaster Metropolitan Borough Council, confirms the impact of the technology on flood planning.