

<b>Institution:</b> University of Southampton		
<b>Unit of Assessment:</b> 11 Computer Science and Informatics		
<b>Title of case study:</b> 11-07 Pioneering transnational open data innovation ecosystems to develop AI and the data economy in a secure way		
<b>Period when the underpinning research was undertaken:</b> 2000 – 2019		
<b>Details of staff conducting the underpinning research from the submitting unit:</b>		
<b>Name(s):</b>	<b>Role(s) (e.g. job title):</b>	<b>Period(s) employed by submitting HEI:</b>
Prof Elena Simperl	Professor	November 2012 – January 2020
Prof Dame Wendy Hall	Regius Professor	January 1984 – present
Prof Sir Nigel Shadbolt	Professor	January 2000 – July 2015
Prof Sir Tim Berners-Lee	Professor	December 2004 – July 2016
Dr Kieron O'Hara	Associate Professor	April 2000 – present
Dr Thanassis Tiropanis	Associate Professor	January 2008 – present
Dr Luis-Daniel Ibáñez	Lecturer	May 2015 – present
Dr Johanna Walker	Research Fellow	January 2017 – present
<b>Period when the claimed impact occurred:</b> February 2015 – December 2020		
<b>Is this case study continued from a case study submitted in 2014?</b> N		

## 1. Summary of the impact

Through the development of novel technologies and socio-technical methods to increase access to and use of data, the University of Southampton's Web and Internet Science Research Group (WAIS) has revolutionised value extraction from data across industry and government. The Group created a framework for transnational, data-driven innovation ecosystems with the design and delivery of two open data business accelerator programmes, **ODINE** (2015-2017) and **Data pitch** (2017-2019), at a Europe-wide level. Over their funded periods, these programmes were evaluated to have created a total of EUR45m (GBP36m) in sales, jobs, contracts investments and efficiencies. Their cumulative economic impact up to the end of 2020 was independently estimated at EUR132m (GBP107m), with the creation of 896 jobs across the UK and Europe. They also supported the development of novel data-centred products and services which have benefited society, the economy and the environment. WAIS research has influenced high-level UK government policy aimed at growing the data economy by shaping the design of new governance frameworks for open, shared and closed data.

## 2. Underpinning research

Data drives innovation and productivity improvements for organisations that collect or control it, but also through its use by other parties. The key to a productive data ecosystem is to enable safe reuse of data and integration with other open, closed and shared datasets in a data value chain. Research within WAIS has sought to optimise value extraction from data across government and industry and, in doing so, has pioneered the development of the open data movement.

Southampton research into data-driven innovation began in 2000 under Shadbolt, Hall and O'Hara, leading to frameworks and technologies for publishing structured and linked data, alongside unstructured content, to enable advanced automation and analysis [3.1]. This was underpinned by interdisciplinary Web Science research into the sociotechnical factors of data use. They created models and methods to augment structured datasets with machine-understandable metadata, to use in large, open, decentralised ecosystems [G1; 3.3]. This enabled the organisation, management, integration and publishing of enriched datasets, applied in several flagship digital government projects, including the UK's open government data portal 'data.gov.uk' [3.3] and the European Data Portal (EDP) [G3]. WAIS research has charted the blockers to effective data use [3.3, 3.6]. For example, studies of government data portals highlighted that: (i) releasing data is just a first step towards wider use; (ii) solving real-world problems requires diverse frameworks and techniques to integrate data sources; and (iii) data holders need support to identify and extract value from their data.

## Impact case study (REF3)

Southampton academics (Hall, Shadbolt, Tiropanis) and colleagues from other universities initiated research into 'Web Observatories' – global resources holding data, visualisations and catalogues [3.2]. As data is not technically open, data holders can collaborate to extract value while still managing the risks of sharing. This research inspired the creation of World Wide Web Consortium (W3C) Community Groups, open online forums where developers and designers can publish and discuss ideas, and institutional frameworks to manage data sharing risk, including data trusts – legal structures that provide independent, fiduciary stewardship of data.

The EDP was launched in 2015 with the publication of 240,000 datasets that could be accessed, and reused by anyone. Simperl extended previous WAIS open data work through Southampton's contribution to the further development of the EDP [G3], undertaking a cross-portal analysis of dataset use across 78 portals in 35 European countries. This work produced guidelines for portal designers to improve user experience and facilitate discovery and sense-making [3.6]. She also proposed novel frameworks to democratise access to and use of data beyond the open data community [3.4] including by citizens, who were recognised as primary drivers of data-driven innovation in [3.3]. In parallel, emerging research by Walker and a PhD student highlighted the need for new forms of engagement with open data to boost innovation. These experiences enabled Southampton researchers (PI: Simperl) to design and deliver two Horizon 2020 open data incubator programmes – ODINE (2015-2017, G2) and Data Pitch (2017-2019, G4) – that facilitated the sharing of commercially valuable closed data owned by multinationals and public institutions with start-ups and SMEs to create new data-driven products and services. A key partner was the UK's Open Data Institute (ODI), which the University of Southampton, under Shadbolt, had co-founded in 2012.

Privacy concerns inhibit data-driven innovation and O'Hara has researched techniques for anonymising data. Anonymisation is technically unsolvable, but O'Hara and colleagues showed that it is possible to manipulate the environment within which the data exists in order to render the risk of reidentification negligible. Their proposed *functional anonymisation* requires a Web Science view beyond the data to its sociotechnical and organisational context; they argue that anonymisation can be a critical part of the toolkit of the privacy-respecting data controller and the wider remit of providing accurate and usable data [3.5].

### 3. References to the research

- 
- 3.1 Shadbolt, Nigel, Berners-Lee, Tim and Hall, Wendy (2006) The Semantic Web Revisited. *IEEE Intelligent Systems*, 21 (3), 96-101. <https://doi.org/10.1109/MIS.2006.62>
  - 3.2 Tiropanis, Thanassis, Hall, Wendy, Shadbolt, Nigel, De Roure, David, Contractor, Noshir and Hendler, Jim (2013) The Web Science Observatory. *IEEE Intelligent Systems*, 28 (2), 100-104. <https://doi.org/10.1109/MIS.2013.50>
  - 3.3 Shadbolt, Nigel, O'Hara, Kieron, Berners-Lee, Tim, Gibbins, Nicholas, Glaser, Hugh, Hall, Wendy and schraefel, m.c. (2012) Linked open government data: lessons from Data.gov.uk. *IEEE Intelligent Systems*, 27(3), Spring Issue, 16-24. <https://doi.org/10.1109/MIS.2012.23>
  - 3.4 Eil, Basil, Harth, Andreas and Simperl, Elena (2014) SPARQL query verbalization for explaining semantic search engine queries. In Presutti, Valentina, d'Amato, Claudia, Gandon, Fabien, d'Aquin, Mathieu, Staab, Steffen and Tordai, Anna (eds.) *The Semantic Web: Trends and Challenges*. Cham, CH. Springer International Publishing, pp. 426-441. [https://doi.org/10.1007/978-3-319-07443-6\\_29](https://doi.org/10.1007/978-3-319-07443-6_29)
  - 3.5 Elliot, Mark, O'Hara, Kieron, Raab, Charles, O'Keefe, Christine M., Mackey, Elaine, Dibben, Chris, Gowans, Heather, Purdam, Kingsley and McCullagh, Karen (2018) Functional anonymisation: personal data and the data environment. *Computer Law & Security Review*. <https://doi.org/10.1016/j.clsr.2018.02.001>
  - 3.6 Koesten, L. M., Kacprzak, E., Tennison, J. F., & Simperl, E. (2017). The Trials and Tribulations of Working with Structured Data: a Study on Information Seeking Behaviour. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 1277-1289). ACM. <https://doi.org/10.1145/3025453.3025838>

### Key grants supporting underpinning research

- G1 PI Shadbolt EPSRC Funding large Grant EnAKTinG the unbounded Web of Data GBP1.94m (2009-2012 EP/G008493/1).

## Impact case study (REF3)

- G2** PI Simperl H2020 Open Data Incubator for Europe (ODINE) EUR7.8m (EUR6.2m to UoS), (2015-2017 grant 644683).
- G3** Co-I Simperl European Data Portal (2015-2018) EUR79,504 (EC contract 30-CE-0677290/00-65), and follow-up funding (2018-21) EUR208,700
- G4** PI Simperl H2020 funded Data Pitch EUR7m (EUR5.9m to UoS), (2017-2019 grant no. 732506).

### 4. Details of the impact

Europe's open data market was worth EUR184bn in 2019, according to consultancy Capgemini Invent. Based on its research expertise in transforming access to, and use of, data across the UK, [3.1, 3.3-3.4, 3.6], Southampton researchers, through GBP12m in funding from 2015 to 2019, have led the design and delivery of two pioneering open data accelerator programmes at a Europe-wide scale. These led directly to increased sales revenues for SMEs, the estimated creation of over 896 new jobs [5.1, 5.6] and a range of societal benefits arising from new products and services. As a recognised thought-leader in data-driven innovation, Southampton has applied its research on data sharing and privacy [3.2, 3.5] to the design of key government policies intended to accelerate the growth of the UK's data economy.

#### The economic and social impact of ODINE

The Southampton-led Open Data Incubator for Europe (ODINE) was a series of six-month programmes for start-ups and SMEs across Europe looking to innovate with open data to develop products and services. The consortium included data facilitators and incubators Fraunhofer IAIS and the ODI, infrastructure provider Telefonica, and media partner *The Guardian*. Over 20 months, the ODINE project provided EUR5.6m (GBP4,200,000) to 57 companies (out of 731 applicants) from 18 European countries. Southampton researchers designed the complex socio-technical framework to facilitate the growth of the ODINE ecosystem and negotiated data sharing agreements between data providers and SMEs. In addition to investment, SMEs received mentoring and incubation support, access to data services and computing infrastructure, and high visibility through articles in *The Guardian*.

An independent assessment [5.1] by consultancy IDC, published in July 2017, concluded that ODINE 'achieved its main objective' to 'accelerate (companies') time to market and chances of success'. It said the ODINE businesses 'contribute(d) to the development of an Open Data ecosystem in Europe covering all segments of the data value chain'. It found that the 57 firms gave ODINE 'high scores in terms of value-added, with the highest benefits concerning accelerating time to market, improving the business idea, and improving the team skills'. Specifically, 97% said that without ODINE 'their time to market would have been longer'. Over the life of the programme itself (2015-2017), ODINE directly created 268 jobs and EUR22.5m (GBP16,500,000) in value through additional jobs, contracts, follow-up investments and efficiency gains [5.2]. A key focus of the IDC report was to measure ODINE's longer-term impact in order to evaluate its sustainability, so it also estimated the economic impact delivered by ODINE up to 2020: EUR110m (GBP88,000,000) of cumulative revenues and the creation of 784 jobs. Average revenues per company by 2020 were estimated to be around EUR1m (GBP890,000), corresponding to EUR55,000 (GBP49,000) in revenues per employee which was described as 'sufficient for sustainability'. This meant that each euro invested by the EC in ODINE would generate up to 14 euros in cumulative revenues by 2020.

The assessment found that ODINE 'succeeded in inspiring and promoting a range of new business ideas', especially in the fields of geospatial mapping and environmental data. The report identified another priority as 'the emerging sustainable or low carbon economy, with several companies focused on energy saving, environmental monitoring, smart mobility'. As an example, Germany-based Green City Solutions used urban air quality data to develop its *CityTree*, a cross between a bench and a hedge that the company says absorbs as much pollution as 275 trees [5.3]. A UK-based company Plume Labs is using open data for hyperlocal air quality forecasts. Since being funded by ODINE in 2015, the company has won a CES Innovation Award (run by the Consumer Technology Association) and it raised USD4.5m (GBP3,600,000) in investment in December 2016 [5.3].

Public authorities have benefited too. For example, the City of Delft (Netherlands) is using data on gritting routes and traffic incidents to optimise its winter de-icing strategy. The City's Information

Advisor said *“We have been able to utilise the insights from ODINE to support our own open data opening and use. ... Our objective is to be more dynamic in their prioritisation and therefore more responsive and relevant to the needs of our citizens. ... We have used insights, best practices and guidance from ... ODINE ... to support the design of the necessary technical processes, and to assess various aspects of governance. For example, research insights on different types of Open Data (clean and raw) illustrated the value to be gained from opening both kinds rather than only publishing clean data, and this is what we have done. This eventually led to greater citizen engagement with our datasets and prompted opening new ones. Further we have shared these learnings and best practices with 12-16 local authorities in the region collaborating on opening their data. We find the insights generated by ODINE to provide a comprehensive blueprint for cities to begin implementing and gaining value from opening and sharing data.”*

Representative examples of how ODINE companies benefited from the programme include [5.3]:

- Idalab (Germany), a data science company that helps cities solve zoning problems, said ODINE *‘allowed us to find product-market fit and bridge the long sales cycles’*.
- OriginTrail (Italy) helps food brands build transparency and authenticity through real-time data on where and how ingredients are sourced. It said ODINE provided *‘the support we needed to start addressing the transparency problem on a global scale’*.
- Tilde (Latvia), a language technology service provider, said ODINE *‘helped us to collect, process and validate massive volumes of multilingual open data in multiple languages ... (which) ... will now enable us to build even more powerful MT systems for clients’*.

Seventy-four articles on the ODINE project were published by *The Guardian*, averaging 500 unique views and 50 shares, with one article on a new open data standard for the banking industry receiving around 80,000 unique pageviews, 500 shares and 300 comments [5.4].

### **The economic and social impact of Data Pitch**

Data Pitch, again led by Southampton, brought together data owners with start-ups to generate fresh ideas for data-driven products and services. In creating this network, Southampton researchers explored the critical factors that were impacting on the way that organisations were creating value from sharing data. They defined and developed 27 data-driven innovation challenges and drew up a legal and privacy toolkit to set the guidelines for data sharing. Large companies that agreed to share their data with start-ups included: Met Office, Konica Minolta, Netherlands telecoms company Altice, and Portuguese healthcare company Jose de Mello Saude. Over two years, 47 start-ups (out of 239 applicants) from 13 countries participated.

An independent evaluation [5.5], compiled by London Economics and published in December 2019, concluded: *‘Data Pitch enabled data-driven innovation that would not otherwise have taken place and had a positive impact on the commercial performance of the participating start-ups.’* It said Data Pitch had *‘succeeded in laying the groundwork for a sustainable open innovation ecosystem in Europe’* and had *‘acted as a demonstrator for data-driven open innovation’*. Over the course of the programme, EUR22.4m (GBP19,700,000) in sales, investments and efficiencies were unlocked by accelerated companies and 112 additional jobs were created [5.6]. Start-ups generated an average of EUR599,432 in sales and EUR338,862 in investment per gigabyte of data shared with them through Data Pitch [5.5]. One year after the 2018 cohort of start-ups had participated in Data Pitch, the ROI on total Data Pitch resources stood at 103% and leveraged investment was 278%, the report said. As an indication of sustainability, the report projected that total annual revenues for all start-ups funded through Data Pitch would increase to EUR35,800,000 in 2022.

The evaluation highlighted representative case studies of how start-ups had benefited:

- Through Data Pitch, Spanish logistics start-up OBBU collaborated with Grenier Packaging, one of Europe’s leading plastic packaging manufacturers; this enabled OBBU to *‘enter a new industry’ which could lead to ‘contracts with other firms in adjacent industries’*.
- Data Pitch enabled German firm Bliq, which offers urban mobility solutions, *‘to pursue a line of development which they would have been unable to do otherwise’*.
- Radiobotics (Denmark) were able to develop their algorithm for advancing musculoskeletal radiology to a stage where they could pitch for investment. This led to additional funding.

### Informing government policy to develop the UK's data economy and AI industry

Based on her Semantic Web and open data research [3.1-3.3], Hall was asked by the Government to conduct a Review on 'Growing the Artificial Intelligence Industry in the UK', alongside Jérôme Pesenti, CEO of BenevolentTech. Published in October 2017 [5.7], the first of the Review's 18 recommendations addressed the need to facilitate data-sharing. It said: 'Government and industry should deliver a programme to develop Data Trusts – proven and trusted frameworks and agreements – to ensure exchanges are secure and mutually beneficial.' In response to this recommendation, Simperl was invited by the Government's Office for AI and the Turing Institute to share experiences of the EDP, ODINE and Data Pitch projects to inform the design of legal and organisational frameworks for sharing closed data. Simperl and Walker were invited to write a White Paper on engagement models for data trusts, presented to the Office for AI, the Government Office for Science, and Defra. A Policy Adviser with the Office of AI said this had 'helped us map out the data sharing landscape and develop our thinking on how to shape the Office for AI's work stream to explore data trusts (a key commitment in the AI Sector Deal) so that it adds to this landscape' [5.8]. From 2019, Hall and O'Hara have convened a working group with the Royal Society and the Law Society of London to produce standards for data trusts. Hall has been named the UK's first AI Skills Champion, and Chair of the Ada Lovelace Institute [5.8]. Finally, a senior advisor on Internet Policy at the Cabinet Office said that "the introduction to Kieron O'Hara has been very fruitful. Our conversations with him have significantly shaped our thinking ahead of the expected Spending Review this year, including products which we're using to challenge departmental thinking and ambition with respect to the geopolitics of internet governance." [5.8]

O'Hara's anonymisation work [3.5] led to the Anonymisation Decision-making Framework, published in 2016 by the UK Anonymisation Network (UKAN), a consortium that includes Southampton, the University of Manchester, ODI and the Office for National Statistics. Freely available, it can be applied to any data where confidentiality is an issue, but sharing is valuable. It was adapted for Australian law in 2017 and for GDPR in 2020. Books describing the techniques have received 16,000 downloads. The Wellcome Trust commended the 'authoritative' Framework to the Department of Health for 'a clear, comprehensive account of what needs to happen to data and the environment in which it is used in order for it to be considered anonymised information' [5.9]. The Medical Research Council also recommended use of the Framework [5.10].

### 5. Sources to corroborate the impact

---

- 5.1 Impact assessment of ODINE by IDC.
- 5.2 Open Data Institute, ODINE deliverable D3.3 Summary of the program, lessons learned, and best practices (2017).
- 5.3 Case studies of ODINE companies: <https://opendataincubator.eu/stories/>. Highlighted Plume's Flow featured in Wired, <https://www.wired.co.uk/article/flow-pollution-tracker>, Green City Solutions' CityTree featured in Wired: <https://www.wired.co.uk/article/citytree-air-pollution-uk-piccadilly>.
- 5.4 Articles published in *The Guardian* on each start-up of the 8 cohorts. Highlighted banking article: <https://www.theguardian.com/media-network/2016/feb/16/banking-industry-uber-moment-standard-open-banking-working-group>
- 5.5 Impact assessment of Data Pitch by London Economics.
- 5.6 Infographic from the Data Pitch homepage: <https://datapitch.eu/> captured on 30-Dec-20. These figures were generated by the ODI to populate the infographic, based on a survey after the London Economics report. These impact numbers were reported to the EC in February 2020. The ODI collected this additional information through running 2 surveys to both cohorts at the beginning of the year. In addition to this, certain start-ups notified the ODI of recent rounds of investment raised.
- 5.7 Growing the Artificial Intelligence Industry in the UK: <https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk>
- 5.8 Statements from UK Gov't Office for AI; Information Advisor, City of Delft; UK Cabinet Office.
- 5.9 Wellcome Trust's response to 'National Data Guardian for Health and Social Care's Review of Data Security, Consent and Opt-Outs', para 16, <https://wellcome.ac.uk/sites/default/files/NDG-review-data-security-consent-Sep16.pdf>
- 5.10 MRC Research Ethics Series 'Using Information About People in Health Research', pp.2, 20, <https://mrc.ukri.org/documents/pdf/using-information-about-people-in-health-research-2017/>.