Impact case study (REF3b)

Institution: Aston University

Unit of Assessment: 11: Computer Science and Informatics

Title of case study: Improving the Management of Uncertainty on the Web: UncertML

1. Summary of the impact

Aston University researchers developed and maintain the Uncertainty Markup Language (UncertML) for quantitative specification and interoperable communication of uncertainty measures in the Web. It is the only complete mechanism for representation of uncertainty in a web context. UncertML has been:

- Used in policy and decision making by UK (Food and Environment Research Agency) and international (European Commission) government agencies, and many research / industrial institutes;

- Presented at industrial /technical workshops, leading to ongoing international collaborations with bodies such as national space agencies (ESA and NASA) and government data providers;

- Accepted as a discussion paper for formal standardisation by the Open Geospatial Consortium;

- Chosen by independent data providers for efficient sharing of complex information and rigorous risk analysis across scientific domains such as pharmacy, global soil mapping and air quality.

2. Underpinning research

UncertML is an underpinning technology enabling the communication and management of quantitative information about error, data quality and uncertainty. It addresses an important challenge in the fast-developing global array of Web-based services and frameworks which collect, discover and exchange data and perform modelling on that data – namely, how to transfer reliably and use the uncertainty information, which is vital for estimating prediction reliability and risk, in an interoperable way so that scientists and policymakers from different domains can properly use it.

During research for the EC-funded INTAMAP project [3.1], Dan Cornford (Reader, at Aston since 1998), Lucy Bastin (Senior Lecturer, at Aston since 2003) and Matthew Williams (PhD/postdoc at Aston, 2006-2012), with initial input from Edzer Pebesma (now at University of Muenster) designed and developed UncertML, a markup language for describing probabilistic uncertainty. UncertML was accepted as a discussion paper in 2009 by the Open Geospatial Consortium [3.5]. UncertML remains the only complete mechanism for representation of uncertainty in a Web context [3.3].

The work led to the EC-funded UncertWeb project, coordinated by Dan Cornford. UncertWeb took forward the concept of interoperability into the "Model Web" domain, where models and data resources exposed as Web services can be composed to produce complex scientific and decision making workflows [3.2, 3.5]. UncertWeb developed tools and technology for managing uncertainty in such a setting, and UncertML is the keystone to the overall UncertWeb framework. The statistical technology employed was informed by the research findings of the MUCM and MUCM2 projects (see section 3). Dan Cornford was the principal MUCM investigator at Aston and was central to the creation of the Web-based MUCM toolkit which complements UncertML-based tools developed in UncertWeb.

UncertML is maintained by Dan Cornford and Lucy Bastin. It has research outcomes and uses as follows (these are later referenced by number in section 4):

Outcome 1: Easy data exchange in the syntactic web. The UncertML convention, its concrete encodings in JSON and XML, and Aston's APIs allow fast and reliable exchange of error and uncertainty data between applications.

Outcome 2: Links to the developing semantic web. UncertML's URI-based vocabulary allows linkage between domains and easy extension of widely-used commercial and open data formats to associate them with well-defined statistical concepts.

Outcome 3: Enabling interoperability for scientific modellers. UncertML as a shared convention allows scientists to leverage the power of re-usable workflows where selected inputs

may be changed for fast and efficient scenario evaluation. It builds on the research findings of pan-European projects with high-profile end-users, and on links outside Europe with NASA, NOAA and Australasian researchers, to ensure relevance of the modelled phenomena and usability of UncertML for a variety of real scientific problems.

Outcome 4: Enabling interdisciplinary research: the modular, domain-agnostic design of UncertML allows use across varied disciplines such as hydrology, biochemistry and ecology, and easy combination with any other Web standards.

Outcome 5: Supporting robust analysis of risk: the explicit management of uncertainty allows its propagation through complex analyses to assess the reliability of predicted results, and clearer communication of uncertain results to the public and other stakeholders.

---

3. **References to the research** (the three that best indicate the quality are marked '*').
Citation counts are from 31/7/2013.

3.1. * Pebesma, E., Cornford, D., Dubois, G., Heuvelink, G. B. M., Hristopoulos, D., Pilz, J., Stöhlker, U., Morin, G. and Skoien, J. O., 2011. INTAMAP: the design and implementation of an interoperable automated interpolation Web service, Computers and Geosciences, 37, 3, 343-352. **(35 citations) DOI:** 10.1016/j.cageo.2010.03.019

3.2. Williams, M., Cornford, D. and Bastin, L., 2008. Describing and Communicating Uncertainty within the Semantic Web, 7th International Semantic Web Conference, October 2008, Karlsruhe, Germany. **(6 citations)   URL:** http://eprints.aston.ac.uk/10038/

3.3. Caumont, H. (Editor) 2010. GEOSS AIP-3 Engineering Report of the Data Harmonisation Working Group, available from http://www.ogcnetwork.net/pub/ogcnetwork/GEOSS/AIP3/pages/AIP-3_ER.html.

3.4. * Williams, M., Cornford, D., Bastin, L. and Pebesma, E., 2009. Uncertainty Markup Language: UncertML, OGC Discussion Paper 08-122r2, Open Geospatial Consortium. **(18 citations) URL:** http://portal.opengeospatial.org/files/?artifact_id=33234

3.5. * Bastin, L., Cornford, D., Jones, R., Heuvelink,G.B.M., Pebesma, E., Stasch, C., Nativi,S., Mazzetti, P. and Williams, M. (2013) Managing Uncertainty in Integrated Environmental Modelling: The UncertWeb framework, Environmental Modelling and Software, 39: 116-134. **(13 citations) DOI:** 10.1016/j.envsoft.2012.02.008

3.6. Yang, K., Blower, J., Bastin, L., Lush, V., Zabala, A., Maso, J., Cornford, D., Diaz, P. & Lumsden, J. (2012)**.** An Integrated View of Data Quality in Earth Observation**.** Philosophical transactions of the Royal Society A*, 371* **(5 citations) DOI:** 10.1098/rsta.2012.0072.

Major grants, including the following, enabled and resulted from this research; they provide evidence of its quality. As described in section 2, Aston expertise was crucial to these projects.

INTAMAP project (INTeroperability and Automated MAPping, http://www.intamap.org): funded by the European Commission under the 6th Framework programme (1.8m Euro) Ran from September 2006 to August 2009, led by Dr Edzer Pebesma then at Utrecht University.

UncertWeb project (Uncertainty in the Model Web - http://www.uncertweb.org). Funded by the European Commission under the 7th Framework programme (2.5m Euro). Ran from February 2010 to January 2013, led by Dr Dan Cornford at Aston University. Tools developed in the project are listed here: https://wiki.aston.ac.uk/foswiki/bin/view/UncertWeb/UncertWebSoftware

MUCM (Managing Uncertainty in Complex Models) and MUCM2 projects (http://mucm.ac.uk). Funded by Research Councils UK (£3M total). Ran from June 2006 to September 2012, led by Prof Tony O'Hagan at Sheffield University. Toolkit available at http://mucm.aston.ac.uk/toolkit.

GeoViQua project (QUALity aware VIsualization for the Global Earth Observation System of systems - http://www.geoviqua.org/). Funded by the European Commission under the 7th Framework programme (4m Euro). Runs from February 2011 to January 2014, led by Dr Joan Maso at the Autonomous University of Barcelona.

---

4. **Details of the impact** (*numbered outcomes are as declared in Section 2. Evidence items (section 5) cited by letter, e.g., 5A, 5B. References (section 3) cited as 3.1, 3.2 etc*)

Our maintenance of UncertML as a free and open standard with excellent documentation, supporting toolkits and version control has facilitated its use by industrial, research and policymaking bodies worldwide to make more informed use of their data in real-world settings. This open approach is well established as the best route to impact for markup languages: their impact depends on widespread adoption by many stakeholders. We have made strategic use of collaborations (e.g. European projects, research networks) to promote international usage of the language and tools. The other key aspect of creating impact is standardisation, so that data can be shared and systems interoperate. This is being actively pursued. In March 2013, UncertML was specifically recommended by the US Library of Congress for digital data collections [5G].

In 2012 the Aston team generated tools for utilising UncertML, including a Web-based elicitation suite and tools for sensitivity analysis and model emulation. The tools are already well adopted: e.g. the elicitation tool alone has over 150 external users across the world, including the Centre for Workforce Intelligence, US Fish & Wildlife Service, California Ocean Science Trust, Swaziland National Trust Commission, UK Health Protection Agency, Sandia National Laboratories (DoE), Australian Customs and Border Protection Service and the Nuclear Decommissioning Authority [5D]. **(Outcomes: 1, 3, 4, 5)**

Policy and decision-making (Outcomes: 1, 3, 4, 5).

The UK Food and Environment Research Agency (FERA) used UncertML and associated Aston tools in 2011-12 to predict food security in the UK. FERA created a workflow to assess the likely impact of climate change and farmer behaviour on wheat production in the UK. This demonstrated that when uncertainties are properly accounted for, the impact of climate change on production is impossible to confidently predict using current data. This important result highlights "the overwhelming risk of policymaking with data of limited quality and the need for focussed improvements in model inputs to predict and respond to future variability in production and its economic effects" (FERA, 2012). This work was presented in April 2012 to the Chief Scientist within the Department for the Environment and Rural Affairs (DEFRA), and their intention is to use similar approaches in policy setting in the future. UncertML specifically contributes here because the workflow can be easily re-run with data pulled from real-time sensor networks or alternative climate models with no loss of statistical detail, allowing multiple scenarios to be tested with much less investment of time and effort. [5A]

UncertML is integral to the eHabitat tool developed at the European Commission's Joint Research Centre (Italy), as part of the Digital Observatory for Protected Areas, which allows stakeholders to monitor and forecast biodiversity, and presented to the UN Convention for Biological Diversity. The Hyderabad meeting of October 2012 recognised these tools as vital in building capacity towards Aichi Biodiversity Target 11, which aims to safeguard inland water, marine and coastal ecosystems. [5B]

Environmental forecasting, monitoring and protection (Outcomes: 1, 3, 5).

The Norwegian Institute for Air Research (NILU) used UncertML and associated technologies in 2012 to develop a prototype probabilistic air quality forecasting system for Oslo, Norway. The system allows NILU to provide uncertainty estimates on forecasts of two key pollutants affecting human health, nitrogen oxides and particulate matter. NILU are "exploring the production of an operational system which will communicate confidence in the forecasts to decision makers (particularly local authorities), who may then better inform the public and plan the management of the transport system" (NILU,June 2013). [5C]

Contribution to international standards (Outcomes: 1, 2, 3, 4).

UncertML was taken to the Open Geospatial Consortium (OGC), the main international geospatial standards body, as a discussion paper [3.4]. The online UncertML vocabulary and schemas are accessed by over 200 unique users per month [5D].

UncertML was adopted within the 2010 GEOSS architectural implementation pilot 3 (AIP-3) as the sole recommended method for harmonising the representation of accuracy [3.3]. GEOSS (Global Earth Observation Systems of Systems) is the global network of content providers allowing decision makers to access information from observing systems such as satellites and sensor networks.

This work led to the GeoViQua project, which incorporates uncertainty into GEOSS and into the

GEO (Group on Earth Observations) portal [5C, 3.6]. GeoViQua's UncertML-based schemas were accessed by 378 unique users last year [5E].

UncertML underlies NetCDF-U, an uncertainty-enabled profile of the popular NetCDF data transmittal standard which was developed at the Institute for Atmospheric Pollution, Italy, presented to the OGC's Earth System Science Domain Working Group and taken up as a discussion paper in November 2011.

UncertML is actively used since mid-2012 in data encodings for other scientific domains: e.g., CellML, PharmML and Systems Biology Markup Language (developed at the European Biological Research Institute) [5F].

International collaboration and data sharing (Outcomes: 1, 2, 3, 4, 5).

UncertML is widely used to define unambiguously uncertainties in Earth observation. In Europe UncertML was used (2010-present) across several major projects including ENVISION, NETMAR, EO2Heaven, EnviroFI and SANY [5H]. These projects represent key developments towards an interoperable geospatial infrastructure and will have significant long term benefits for the ways we observe and model our environment.

In Australasia, UncertML is used by government institute Landcare Research New Zealand in their GSMML data format, proposed as the international standard for the Global Soil Map [5I] which aggregates data from 194 countries to generate a vital resource for 'better decisions in global issues like food production, hunger eradication, climate change, and environmental degradation'.

5. Sources to corroborate the impact

A. *Contact for evidence of FERA impact:* Geographical Information Scientist, The Food and Environment Research Agency, UK.

B. In the COP11 meeting of 2012 (Hyderabad, India) one of the 33 decisions taken specifically references the DOPA tools, which rely on UncertML for handling of uncertain data, as vital in building capacity towards Aichi Biodiversity Target 11: http://www.cbd.int/doc/decisions/COP-11/cop-11-dec-24-en.doc, section 8.

C. *Contact for evidence of NILU impact:* Senior Scientist, Norwegian Institute for Air Research (NILU), Norway.

D. http://uncertml.org/stats/index.html Access statistics demonstrating that UncertML vocabulary and schema implementations at http://www.uncertml.org/ are regularly accessed by over 140 unique users per month. A list of registered elicitator users is also supplied.

E. http://schemas.geoviqua.org/webstats/ Access statistics demonstrating that GeoViQua's UncertML-based schemas were accessed by 378 unique users in the year to 31/7/2013.

F. SBML (Systems Biology Markup Language) is very widely used in biological modelling and has adopted UncertML. *Contact for evidence of SBML and PharmML impact:* Biomodels Team, European Bioinformatics Institute (EMBL-EBI), UK.The CellML community also aims to use UncertML – see the Introduction / Future Work sections of this 2012 PLOS paper: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0039721

G. http://www.digitalpreservation.gov/formats/content/gis_intro.shtml Sustainability of Digital Formats Planning policy document recommending that UncertML be used alongside established global standards such as FGDC and ISO in maintaining data collections for the US Library of Congress.

H. http://www.envirofi.eu/Portals/89/Docs/Project/Public_deliverables/ENVIROFI%20D5.3_ENVIROFI_data_and_meta_information_specifications.pdf (pages 28, 33, 37, 40). How the EnviroFI project uses UncertML to enable the 'Future Internet'. *Contact for evidence of NETMAR impact* –Senior Earth Observation Scientist, Plymouth Marine Laboratory, UK.

I. http://adsabs.harvard.edu/abs/2013EGUGA..15.6761R. Landcare Research New Zealand's GSMML profile, in which UncertML is a key component. This profile is proposed as the national standard for the Global Soil Map. *Contact for evidence of Landcare impact*: Environmental Information & System Analyst, Landcare Research, New Zealand.