Data Integrity | DOI:10.1145/3589154

The Veracity Grand Challenge in Computing:

A Perspective from Aotearoa New Zealand

BY MARKUS LUCZAK-ROESCH, MATTHIAS GALSTER, AND KEVIN SHEDLOCK

THE NEW ZEALAND government identified numerous challenges related to trust and truth in the context of digital technologies. These challenges result from an ever-increasing amount of online social networks, end-to-end digital supply chains, automated decision-making tools, generative artificial intelligence (AI), and cyber-physical systems. Such challenges impact people's lives across professional and private contexts and led to the *Veracity Project*[a] 2021–2024.

But what is *veracity*? Outside the field of computing, veracity is not a common term in everyday language. One dictionary definition is "conformity with truth or fact."[b] But does that definition really capture the entire meaning of veracity within computing?

Veracity was initially introduced as the fourth "V" to Big Data's original three Vs—volume, velocity and variety—often interpreted as data quality.[3] However, dictionaries also define veracity as the "power of conveying or perceiving truth," highlighting a sender-receiver perspective.

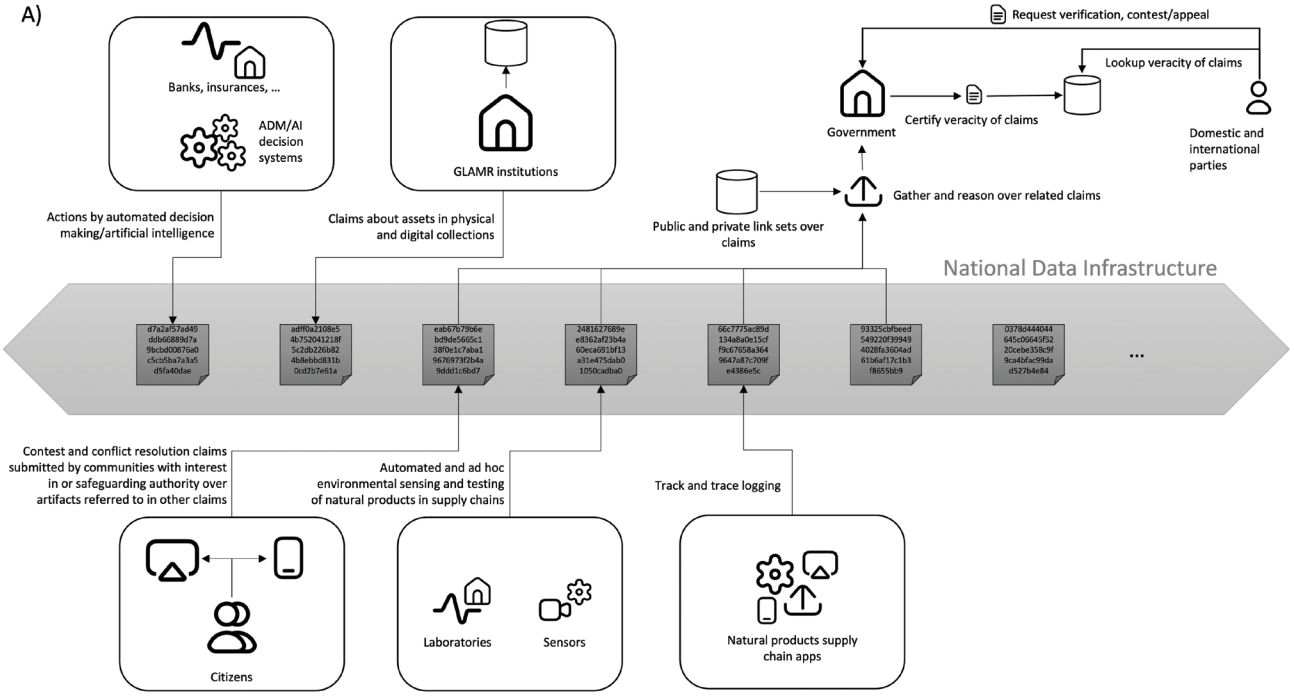
IMAGE BY MASTOCKER

Conceptualizing veracity as data quality does not capture its scope multidirectionally; and many low-trust scenarios arise from someone accessing high-quality data but doing something with that data that harms others. An example is the Facebook-Cambridge Analytica scandal, a low-veracity situation involving high-quality data.

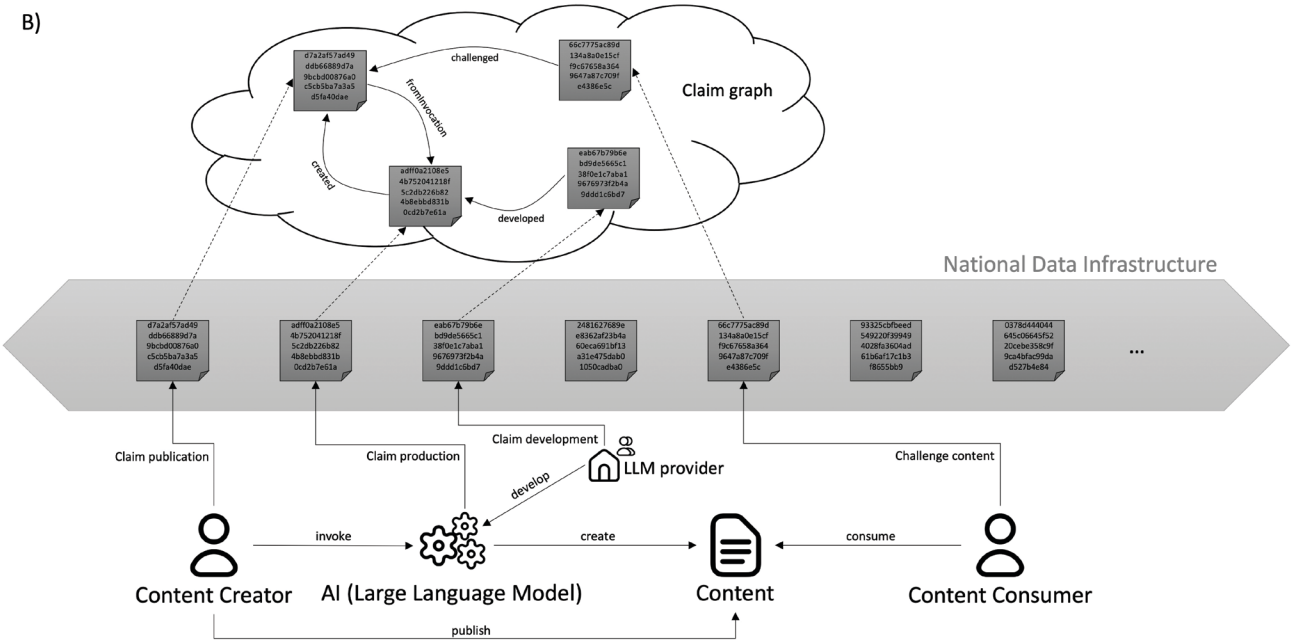
Transcending boundaries between physical and digital spheres. This conceptual ambiguity makes veracity uniquely interesting within Aotearoa New Zealand and other countries where multidirectional considerations around sovereignty of artifacts (including digital artifacts) are culturally and legally embedded in society. New Zealand is an example of a bicultural society with a commitment to recognize language and culture of the Indigenous peoples, Māori. Technology, and how technology is developed and used, must be culturally responsive and inclusive. This is specifically visible in requirements around data sovereignty. Research from Aotearoa New Zealand can move beyond common thinking about computation by fostering a co-design process that allows Indigenous and global perspectives to jointly create this new conceptualization of veracity.

Led by the affiliated Indigenous researchers, the Veracity Project research began from the notion of "Ko te taiao matihiko," emphasizing a holistic view to the world around us, and dissolving boundaries between the physical and digital worlds. This provides a novel perspective on computing in which digital representations of artifacts cannot be disentangled from their grounding in the physical world. This grounding could be either the real-world context of a digital artifact[5] (for example, the person that used a digital service is connected to the data, and the place in which the app was used) or the real-world artifact itself about which data is created (for example, an apple

a https://veracity.wgtn.ac.nz/
b https://www.merriam-webster.com/dictionary/veracity



A) To ensure and prove veracity, people, organizations (for example, from the GLAMR sector including galleries, libraries, archives, museums and record-keeping institutions), and applications (including automatic decision making/ADM systems) preregister secure identifiers (for example, embodying time and digital signatures) related to claims they want to make while maintaining full control over the actual data related to the claim. Claims can be about a certain transaction like the exchange of an artifact between parties, or about the condition or properties of an artifact at a certain time. Witness accounts about the truthfulness of claims as well as contesting of claims can also be submitted. All claims are linked by context and can be examined to assess the state of veracity.



B) Example instantiation of the architectural concept in the context of AI-supported content authoring involving a Large Language Model (LLM). Each actor in the process (including the AI system) submits claims related to their involvement in the content production, publishing and consumption process. A content consumer, for example, may submit a claim that is a challenge against the veracity of the produced content (for example, to flag misinformation or cultural appropriation). A claim graph can be constructed from all claims to infer the veracity of the content publication life cycle and each actors' role and stake.

Icons in figure from Slim Icon Set, MIT License; https://iconduck.com/sets/slim-icon-set.

Conceptual view of the envisioned data infrastructure approach to veracity.

traveling through a supply chain along with supporting metadata). The traditional approach in computing is to introduce digital twins as a "copy" of the real artifacts. But when entangling digital and physical representations, both exist and evolve as one.

This holistic concept of veracity is challenged by blind spots between the physical and digital spheres. Currently, there is no technology available that allows objects of arbitrary makeup (for example, a biological object like an apple) or scale (a car's individual screws) to hold data about itself, its life cycle, or the integrity placed on it by the means of production (organic) without attaching something to it (physical label, chemical nanodots). The disembodiment of data from its origin has become a pressing sociocultural, legal, and technical concern, and makes veracity one of the grand challenges in computing.

Trust, truth, authenticity, and demonstrability. The Veracity Project is investigating cultural and formal approaches to both clarify and explore veracity. Formal approaches help clarify because they abstract away from any example to get a clear view of problems, and then allow expression of that view in a rigorous language, which is equipped with rules that preserve properties (for example, trust, truth, authenticity, and demonstrability). For example, event calculus helps model supply chains, use an "engine" to explore the models to see how robust or complete the real supply chains might be, and develop a logic based on constructive logics[2] to pin down the essence of veracity.

Achieving high veracity through national data infrastructures. There are numerous situations where one party could disregard another party's interests to maximize their own benefit, thus lowering veracity. The Veracity Project is exploring use cases as diverse as tracking the use and integrity of Traditional Knowledge Labels attached to cultural heritage records;[c] supporting transparency and trust in the supply chain of organics products; and making AI decisions contestable. Stakeholder interviews indicate that veracity decomposes (at least) into technical veracity, process veracity, financial veracity, regulatory veracity, and cultural veracity.

Commonly recommended approaches like blockchain do not always support veracity as envisioned here. Veracity may involve real-world interactions; blockchain-based solutions may not capture all the information to establish veracity around those or the principles inherent in blockchains may exclude certain users. Furthermore, blockchain may help ensure the integrity of data records once they are in the system, but we still need mechanisms to ensure data records that go into a system are trustworthy. Finally, some types of blockchain may not be practical in all cases, for example, when enhancing existing systems with provenance mechanisms, or in domains where scalability and energy consumption are key drivers.

The Veracity Project works toward a novel kind of national data infrastructure involving decentralized data management following the principles of 'solid' data pods,[4] pre-registration of secure footprints of digital actions an entity (for example, a person, organization, or artificial agent) wants to be able to make claims or withstand scrutiny when challenged, and the ability for witnesses to vouch for or contest other entities' claims or actions. Related views have recently been articulated about privacy-preserving computation[1] and personal data management.[5]

c https://www.localcontexts.org

The disembodiment of data from its origin has become a pressing sociocultural, legal, and technical concern, and makes veracity one of the grand challenges in computing.

The accompanying figure depicts our concept of such a national data infrastructure that focuses on open standards and interoperability at the technical level, and data ownership, consent, and contesting at the governance level. We investigate which open standards (for example, decentralized identifiers, verifiable credentials) enable infrastructures that scale, keep costs low while supporting legacy technologies, respect data sovereignty, and reduce technical dependencies.

Acknowledgments. The following team members of the Veracity Project also contributed substantially to the concepts and ideas conveyed in this article (listed in alphabetical order): Kelly Blincoe, Stephen Cranefield, Jens Dietrich, David Eyers, Brendan Hoare, Maui Hudson, Tim Miller, and Steve Reeves.

This work is supported under the New Zealand National Science Challenge Science for Technological Innovation (SfTI) spearhead project "Veracity Technology."

References

1. Agrawal, N., Binns, R., Van Kleek, M., Laine, K. and Shadbolt, N. Exploring design and governance challenges in the development of privacy-preserving computation. In *Proceedings of the ACM CHI Conf. Human Factors in Computing Systems*, May 2021, 1–13.
2. Bridges, D. and Reeves, S. Constructive mathematics in theory and programming practice. *Philosophia Mathematica 7*, 1 (1999), 65–104.
3. Rubin, V. and Lukoianova, T. Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online 24*, 1 (2013), 4.
4. Sambra, A.V. et al. Solid: A platform for decentralized social applications based on linked data. MIT CSAIL & Qatar Computing Research Institute, Tech. Rep., 2016.
5. Shedlock, K. and Vos, M. A conceptual model of Indigenous knowledge applied to the construction of the IT artefact. In *Proceedings of the 31st Annual CITRENZ Conf.* July 2018, 1–7.
6. Verbrugge, S., Vannieuwenborg, F., Van der Wee, M., Colle, D., Taelman, R., and Verborgh, R. Towards a personal data vault society: an interplay between technological and business perspectives. In *Proceedings of the 60th FITCE Communication Days Congress for ICT Professionals: Industrial Data–Cloud, Low Latency and Privacy.* IEEE, Sept. 2021, 1–6.

Markus Luczak-Roesch is an associate professor of information systems in the School of Information Management at Victoria University of Wellington, New Zealand.

Matthias Galster is an associate professor of computer science and engineering at the University of Canterbury, Christchurch, New Zealand.

Kevin Shedlock (Ngāpuhi, Ngāti Porou, Whakatōhea) is an assistant lecturer in the School of Engineering and Computer Science at Victoria University of Wellington, New Zealand.

This work is licensed under a https://creativecommons.org/licenses/by-sa/4.0/