

AIDGO: A FAIR Ontology for Aligning EU AI Act Data Governance Obligations with ISO/IEC Technical Standards

Kuruvilla George Aiyankovil^{1,*}, David Lewis^{1,*}

¹Trinity College Dublin, Ireland

Abstract

The EU AI Act (Regulation 2024/1689) introduces legally binding data governance obligations for providers of high-risk AI systems under Article 10. Translating these obligations into actionable technical guidance requires mapping them onto established standards, yet existing crosswalks are informal, undocumented, and not machine-readable. We present AIDGO (AI Data Governance Ontology), an OWL 2 ontology that formalises the alignment between EU AI Act Article 10 and ISO/IEC 5259-3:2025 (data quality management for machine learning). AIDGO introduces a five-type alignment taxonomy grounded in W3C SKOS, ISO 25964-2:2013 and SSSOM, decomposes both instruments to atomic requirement level using the TAIR/OSLC-RM framework, and serialises 24 alignment assertions as PROV-O-annotated RDF queryable by SPARQL. Evaluation against eight competency questions shows that AIDGO is reproducible, FAIR-compliant and extensible. Of the 17 atomic Article 10 obligations, 2 are completely satisfied by ISO 5259-3, 12 are partially satisfied with documented residual obligations, and 3 have no aligned ISO requirement and require legal intervention beyond any technical standard.

Keywords

ontology engineering, EU AI Act, ISO 5259, legal-technical alignment, knowledge graph, SPARQL, SKOS, data governance, FAIR, compliance

1. Introduction

The EU AI Act (Regulation 2024/1689) [1], in force since 1 August 2024, is the first comprehensive, legally binding framework for artificial intelligence. Article 10 (“Data and Data Governance”) imposes technically demanding obligations on providers of high-risk AI systems concerning training, validation and testing datasets, spanning eight governance dimensions: governance practices (Art.10(2)); design choices (a); collection and origin (b); preparation (c); assumptions about what data measures (d); availability and suitability (e); bias examination (f); and identification of data gaps (g)–(h). From August 2026, non-compliance can attract fines of up to €35 million or 7% of global annual turnover. Yet the Act provides no technical methodology for demonstrating compliance, delegating the detail to harmonised standards [2].

ISO/IEC 5259-3:2025 [3], the management-requirements part of the Data Quality for Machine Learning series, is the strongest candidate technical standard for Article 10. It specifies normative requirements across the data life cycle, from motivation and specification through acquisition, pre-processing, augmentation, provisioning and decommissioning. However, no *reproducible*, machine-readable, FAIR mapping between ISO/IEC 5259-3 and Article 10 currently exists. What exists are informal crosswalk spreadsheets that map ISO clauses to articles at coarse granularity. Such mappings are *not reproducible* (no formal criteria), *not machine-readable* (not SPARQL-queryable), *not FAIR* (no provenance or versioning), *not granular enough* (clause-level mapping obscures heterogeneity: Art.10(2) alone decomposes into eight obligations with very different alignment profiles) and *not openly licensed* (typically circulated as internal or consultancy artefacts, so they cannot be independently checked, corrected or built upon). Licensing is not incidental here. Article 10 is public law, freely readable by anyone it binds, whereas ISO/IEC 5259-3 is copyrighted and sold; an alignment resource that is itself closed compounds that asymmetry, leaving the question of what conformance actually buys answerable only by those who

Posters, Demos, Blue Sky, and Tutorials at SEMANTiCS 2026, Sep 2026, Ghent, Belgium

*Corresponding author.

✉ georgeak@tcd.ie (K. G. Aiyankovil); dave.lewis@tcd.ie (D. Lewis)

ORCID 0000-0002-0230-3959 (K. G. Aiyankovil); 0000-0002-3503-4644 (D. Lewis)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

can pay twice. AIDGO therefore publishes its alignment layer — requirement identifiers, alignment types, alignment bases and PROV-O provenance — under CC BY 4.0, reproducing no normative ISO text. Correspondingly, AIDGO claims no authoritative interpretation of Article 10: authority rests with the Commission, the harmonised standards to be developed under Article 40 and ultimately the Court of Justice. Its contribution is of a different kind — an open, inspectable and contestable mapping that can be versioned and superseded as authoritative guidance emerges.

Motivating scenario. A firm implementing ISO 5259-3 asks whether it satisfies Art.10(2)(f) — examine datasets for biases affecting fundamental rights or discrimination. Reading ISO §7.3.5.2 (monitor and address data bias), a compliance officer might conclude “yes”. This is incorrect: ISO treats bias as a data-quality dimension, whereas Art.10(2)(f) frames it as a fundamental-rights protection linked to anti-discrimination law — structurally outside any data-quality standard. AIDGO makes this explicit. It classifies the relationship as `PartiallySatisfies` with basis `ScopeOverlap`, and records the *residual obligation* — the portion of the legal obligation that remains undischarged once the aligned ISO requirement has been implemented. Residual obligations are expressed by `aidgo:alignmentGap` (`owl:DatatypeProperty`, range `rdf:langString`), asserted via RDF-star annotation on the mapping triple alongside `aidgo:alignmentType`, `aidgo:alignmentBasis` and PROV-O provenance; `aidgo:recommendedStandard` names any instrument that discharges the residue. The property is mandatory for every `PartiallySatisfies` and `NoAlignment` assertion, making residual obligations enumerable by SPARQL rather than buried in prose (CQ-M2, CQ-M3).

Contributions. (1) A five-type alignment taxonomy — `CompletelySatisfies`, `PartiallySatisfies`, `RelatedTo`, `ConflictsWith`, `NoAlignment` — grounded in W3C SKOS [4], ISO 25964-2:2013 [5] and SSSOM [6]; (2) a separable mapping layer in which every assertion carries a mandatory controlled-vocabulary basis, and every incomplete satisfaction records its residual obligation, so partial alignments remain extensible; (3) a worked application to Article 10 and ISO/IEC 5259-3 — 17 legal obligations and 190 technical requirements decomposed with TAIR [7] and OSLC-RM [8], yielding 24 alignment assertions validated against eight competency questions; and (4) a public interactive knowledge graph explorer (332 nodes, 346 edges) and full documentation, licensed CC BY 4.0 and published at <https://w3id.org/aidgo>.

2. State of the Art

Legal knowledge representation. LKIF Core [9] showed that OWL 2 is expressive and decidable enough for legal knowledge, a finding that informs AIDGO’s design. Akoma Ntoso [10] established a hierarchical document model that is the structural precedent for AIDGO’s atomic decomposition. PrOnto [11] formalises GDPR concepts for SPARQL-based compliance checking, and the W3C Data Privacy Vocabulary [12] provides the vocabulary-pattern design that AIDGO adopts for its `alignmentBasis` controlled vocabulary. ISO 25964-2:2013 [5] contributes the five-category mapping taxonomy, while SSSOM [6] contributes the mandatory-justification principle that is AIDGO’s most consequential design constraint.

Compliance and AI governance. Francesconi and Governatori [13] formalise compliance checking over linked open data in OWL 2, and Pandit et al. [14] show that PROV-O provenance can satisfy GDPR accountability, an approach AIDGO inherits. AIRO [15] is the most direct structural precursor, representing cross-document relationships between regulatory and technical concepts in SPARQL-queryable form. The TAIR framework [7] is AIDGO’s direct methodological predecessor, supplying the `tair:Requirement` and `tair:ConjunctiveRequirement` classes and the atomic decomposition method that AIDGO adopts rather than contributes. These predecessors are FAIR-compliant; FAIR is therefore a baseline here, not a differentiator. AIDGO’s addition sits at the level of the individual mapping assertion: each correspondence carries a controlled-vocabulary `aidgo:alignmentBasis` recording *why* the interpretive judgement was made, following SSSOM’s mandatory-justification principle [6], and

Table 1

Capability comparison on criteria independent of AIDGO’s design choices.

Work	Mapping locus	Partial corr.	Residual recorded
PrOnto [11]	Single instrument	n/a	n/a
Francesconi 2022 [13]	Embedded in model	No	No
AIRO [15]	Embedded in model	Implicit	No
TAIR [7]	Separate layer	Yes	No
AIDGO (this paper)	Separate layer	Yes	Yes

Table 2

AIDGO class hierarchy and knowledge graph statistics.

Class	Count	Notes
<code>aidgo:LegalRequirement</code>	17	Art.10 atomic obligations
<code>tair:Requirement</code>	71	ISO 5259-3 clause-level reqs
<code>tair:ConjunctiveRequirement</code>	119	Atomic bracket-context sub-reqs
<code>aidgo:RequirementCollection</code>	51	ISO clause groupings
<code>aidgo:Concept</code>	74	ISO 5259 terms, Parts 1–5
<code>aidgo:AlignmentAssertion</code>	24	Legal-technical instances
<code>aidgo:AlignmentType</code>	5	One subclass per type
Total nodes	332	346 edges, 7 edge types

— where satisfaction is incomplete — an `aidgo:alignmentGap` recording the undischarged remainder. Table 1 compares these approaches on capability criteria independent of those choices. *Mapping locus* asks whether correspondences sit inside a source model or in a separate layer, determining whether a mapping can be revised without republishing either instrument — a live concern while harmonised standards under Article 40 remain in development. *Partial correspondence* asks whether a relation can be asserted as incomplete rather than forced binary. *Residual recording* asks whether the undischarged remainder is itself retrievable, which is what lets a partial mapping be extended.

3. The AIDGO Ontology

AIDGO is built on four principles, two inherited and two specific to the mapping layer. Inherited: *atomic granularity* (from TAIR [7], both instruments decomposed to the smallest independently testable unit) and *FAIR compliance* [16] (globally unique IRIs, standards-compliant RDF/Turtle, CC BY 4.0), both shared with AIDGO’s predecessors. Specific to the mapping layer: *mandatory justification* (no alignment assertion may exist without a human-readable rationale, a controlled-vocabulary basis, and complete PROV-O provenance [17], so every interpretive judgement is inspectable and contestable) and *residual modelling* (every incomplete satisfaction records what remains undischarged, making partial mappings extensible rather than terminal). It defines seven top-level classes (Table 2); the 74 concepts span all five parts of ISO/IEC 5259.

The five alignment types are AIDGO’s core ontological contribution. Each is an `owl:Class` subclass of `aidgo:AlignmentType`; the taxonomy is exhaustive and mutually exclusive (Table 3). Every technical requirement additionally carries an `aidgo:dqmlcStage` annotation drawn from the ISO/IEC 5259-3 Data Quality Management Life Cycle, supporting both candidate identification and coverage analysis.

4. Alignment Mapping Methodology

The methodology follows four sequential phases, each grounded in a published specification. It is designed to be replicable: two analysts following the same process on the same text pair should reach

Table 3

AIDGO alignment type taxonomy with SKOS and ISO 25964-2 grounding.

AIDGO type	SKOS equiv.	ISO 25964-2	Criterion
CompletelySatisfies	skos:exactMatch	Exact equiv.	All semantic content addressed; no residual obligation
PartiallySatisfies	skos:closeMatch	Partial equiv.	Same required action; ≥ 1 documented gap
RelatedTo	skos:relatedMatch	Associative	Semantically related; no direct normative satisfaction
ConflictsWith	(none)	(incompatible)	Joint satisfaction impossible
NoAlignment	(none)	(gap)	Standard entirely silent; compliance gap

the same classification, and where they disagree the documented justification makes the disagreement explicit and resolvable.

Phase 1 – Atomic decomposition (OSLC-RM/TAIR). Both instruments are decomposed to their smallest independently testable unit. On the legal side, each sub-clause with a distinct normative obligation becomes one `aidgo:LegalRequirement`, yielding 17 atomic requirements for Art.10(1)–(6). On the technical side, each `tair:ConjunctiveRequirement` is one atomic SHALL statement using the bracket-context pattern:

```
AidgoConjReq_7.3.5.2_Rd:
  skos:definition = "[The organisation shall implement data
  pre-processing measures. The organisation shall:]
  monitor and address data bias and concept drift in
  training data."
```

Without the bracket context the `ConjReq` is ambiguous; with it, a compliance assessor can evaluate the obligation independently. This produces 71 clause-level requirements and 119 atomic `ConjReqs`; mapping is always performed at the `ConjReq` level on the ISO side.

Phase 2 – Candidate identification (ISO 25964-2 §14). Candidates are identified by three complementary methods: terminological matching (`prefLabel/altLabel`), conceptual matching (`definition`), and structural matching (`dqmlcStage` filtering), typically reducing the set to 3–12 per legal requirement.

Phase 3 – Semantic comparison. Each candidate pair is evaluated by a decision sequence applied *in order*; the first condition satisfied determines the type. Each test is stated over the normative content of the two texts – the actions and outcomes they require – rather than over authorial intent, which is not reliably recoverable. (1) Does the ISO requirement address the same regulated subject matter? If not → `NoAlignment`. (2) Does conformance with the ISO requirement entail non-conformance with the legal obligation, so that both cannot be satisfied together? If so → `ConflictsWith`. (3) Does ISO require every action and outcome the obligation requires, at equivalent scope, actor framing and specificity? If so → `CompletelySatisfies`. (4) Does ISO require the same action or outcome but differ on at least one of: narrower scope, missing sub-elements, different actor framing, or lower specificity? If so → `PartiallySatisfies`, with the differing element recorded as the residual obligation. (5) Otherwise a semantic relation exists without normative satisfaction → `RelatedTo`. *Restriction* – ISO applying a requirement to a narrower class of cases – is handled at step 4 as a scope gap, not a conflict: it leaves the obligation under-satisfied, not impossible to satisfy.

Phase 4 – Justification and provenance (SSSOM/PROV-O). Every assertion is recorded with full mandatory metadata. The `aidgo:alignmentBasis` property uses a seven-term controlled vocabulary (`SemanticEquivalence`, `FunctionalEquivalence`, `ScopeOverlap`, `ConceptualCoverage`, `TerminologicalMatch`, `IncompatibleAssumption`, `GapIdentified`). `PROV-O` metadata (`wasGeneratedBy`, `generatedAtTime`, `wasAttributedTo`) is recorded on every assertion, making the mapping set a versioned, citable artefact.

Table 4

Competency question evaluation. All 8 CQs return results consistent with the mapping.

CQ	Question	Result
CQ-M1	Which ISO reqs completely satisfy Art.10 obligations?	6 rows
CQ-M2	Which partially satisfy, with gap evidence?	15 rows
CQ-M3	Which obligations have no ISO coverage?	3 rows
CQ-M4	Any conflicts between ISO and legal obligations?	0 rows (correct)
CQ-M5	Distribution by alignment type (assertions)?	CS=6, PS=15, NA=3
CQ-M6	Provenance of each assertion?	24 rows, full PROV-O
CQ-M7	Alignment distribution across ISO sections?	8 rows
CQ-M8	Terminology-standard contribution?	74 concepts

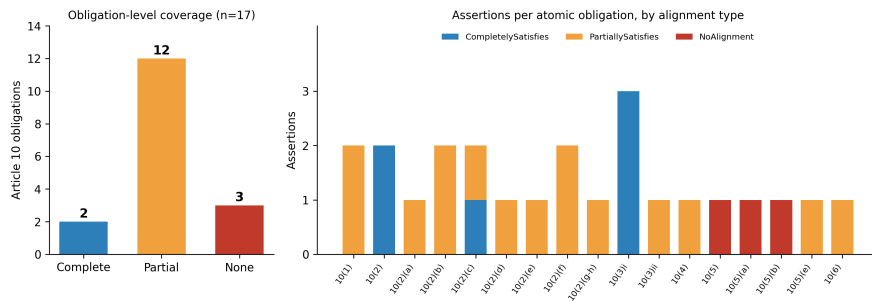


Figure 1: Left: obligation-level coverage over the 17 atomic Article 10 obligations, classified by weakest alignment. Right: assertion count for each atomic obligation, stacked by alignment type, showing that multi-assertion obligations concentrate the *CompletelySatisfies* results.

Worked example (*PartiallySatisfies*). For the motivating case — Art.10(2)(f) vs. ISO §7.3.5.2(d) — steps 1 and 2 pass through (same subject matter, no incompatibility), step 3 fails (the legal text adds a fundamental-rights and anti-discrimination framing outside ISO’s scope) and step 4 matches on narrower scope (the required action, examining and addressing bias in training data, is present), giving *PartiallySatisfies* | basis *ScopeOverlap* with the documented residual obligation recorded as the gap.

5. Results and Evaluation

All 24 assertions are serialised in `aidgo-mappings.ttl` (~40 KB) and evaluated against eight Mapping Competency Questions, each answerable by a SPARQL query (Table 4). Results are reported at two levels, which must not be conflated. At *assertion* level the 24 mappings comprise six *CompletelySatisfies*, fifteen *PartiallySatisfies* and three *NoAlignment*, with zero conflicts. At *obligation* level — the level at which compliance claims are made — the 17 atomic obligations resolve to 2 completely satisfied, 12 partially satisfied and 3 with no aligned ISO requirement (Figure 1). The two levels differ because several ISO requirements may bear on one obligation: five of the six *CompletelySatisfies* assertions fall on just two obligations. An obligation is therefore classified by the *weakest* alignment among its assertions, so that an obligation counts as completely satisfied only if every assertion against it is *CompletelySatisfies*; Art.10(2)(c) is partially satisfied on this rule despite carrying one *CompletelySatisfies* assertion. Counts are reported rather than percentages: with $n = 17$ a single reclassification shifts any proportion by six points.

The six *CompletelySatisfies* assertions fall on only three obligations: data governance practices (Art.10(2), by ISO §6.3.2 and ConjReq §7.3.1.1), data augmentation (Art.10(2)(c), by §7.3.6.1) and dataset quality (Art.10(3), by §7.3.2.3, §7.3.5.2 and ConjReq §7.3.1.1). This concentration requires explanation, since no two ISO requirements can each be *semantically equivalent* to one obligation while differing from

each other. It has two causes. Art.10(3) is *conjunctive rather than atomic*: its text requires datasets to be relevant, representative, free of errors and complete, and the three ISO requirements discharge different conjuncts — §7.3.2.3 representativeness, §7.3.5.2 error-freeness, ConjReq §7.3.1.1 relevance to intended purpose. Phase 1 decomposition did not run deep enough here. Art.10(2), by contrast, is atomic but broadly worded, so two ISO requirements independently address it; this is redundant coverage, for which SemanticEquivalence is too strong a basis. The fifteen partial alignments exhibit four recurring gap patterns: sector-neutrality, rights framing (Art.10(2)(f)), actor specificity (ISO’s “organisation” vs. the Act’s “providers of high-risk AI systems”) and deployment-context specificity (Art.10(4)). All three NoAlignment assertions arise from Art.10(5)’s provisions for processing special categories of personal data — GDPR Article 9 legal-basis requirements that no data-quality standard can satisfy. Art.10(5) is not uniformly unaligned: one sub-obligation is partial. The Data Provisioning stage (§7.3.7) yields no Article 10 alignments, serving Arts. 12 and 15. AIDGO is FAIR by construction: persistent IRIs under the <https://w3id.org/aidgo> namespace (Findable), RDF/Turtle under CC BY 4.0 (Accessible), W3C-standard predicates compatible with Jena, GraphDB, Stardog and Oxigraph (Interoperable), and full PROV-O metadata with owl:versionInfo (Reusable).

6. Discussion and Conclusion

AIDGO gives the first SPARQL-queryable account of how ISO/IEC 5259-3 relates to Article 10, stated at obligation level: of 17 atomic obligations, 2 are completely satisfied, 12 partially, and 3 not at all. ISO 5259-3 conformance therefore cannot be a proxy for Article 10 compliance. The three unaligned obligations, all within Art.10(5), require direct GDPR compliance. The twelve partially satisfied obligations mean organisations have a technical foundation but must close documented residual obligations, notably the rights-framing gap (Art.10(2)(f)) and the deployment-context gap (Art.10(4)). The recurring sector-neutrality gap reflects a structural feature of ISO standards (universal applicability vs. risk-targeted regulation) that cannot be closed by ISO 5259 alone; AIDGO’s machine-readable gap documentation is exactly the input standards developers (e.g. CEN-CENELEC JTC 21) need to plan such work. ConjReq-level mapping is the most consequential design decision: clause-level mapping would report Art.10(2) as merely “partially addressed”, hiding that some sub-obligations are fully satisfied while others are genuine gaps.

Limitations. The mapping reflects the current texts of Article 10 and ISO/IEC 5259-3:2025; forthcoming implementing acts may alter classifications, and AIDGO’s versioning model accommodates such updates. The classifications represent expert judgement, and the methodology is designed for expert-driven manual mapping — automated LLM classification is not yet accurate enough for regulatory purposes, where a misclassification could carry legal consequences. Independent replication would strengthen confidence, especially for borderline partial alignments. Art.10(3) also shows that decomposition did not reach full atomicity everywhere; correcting this raises the obligation count above 17 without altering which obligations are completely, partially or not satisfied.

In summary, AIDGO makes the Article 10 / ISO 5259-3 compliance landscape machine-readable, SPARQL-queryable and FAIR-compliant, through a grounded five-type taxonomy, a four-phase methodology, atomic decomposition and a PROV-O provenance model. The RDF graph, documentation, knowledge graph explorer and mapping workbook are available under CC BY 4.0 at <https://w3id.org/aidgo>.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Grammarly for grammar and spelling checks, and GitHub Copilot to help format the RDF/Turtle serialisations and SPARQL queries. The authors reviewed and edited all content and take full responsibility for it.

References

- [1] European Parliament and Council, Regulation (eu) 2024/1689 laying down harmonised rules on artificial intelligence (ai act), 2024. Official Journal of the European Union, L 2024/1689, CELEX:32024R1689.
- [2] H. Junklewitz, R. Hamon, T. Evas, A. André, S. Scalzo, Analysis of the preliminary AI standardisation work to support the implementation of the proposed AI Act, Technical Report JRC Technical Report EUR 31091 EN, European Commission Joint Research Centre, 2023.
- [3] ISO/IEC, ISO/IEC 5259-3:2025 data quality for analytics and machine learning (DQML)—part 3: Data quality management requirements, 2025. Geneva: ISO.
- [4] A. Miles, S. Bechhofer, SKOS simple knowledge organization system reference, W3C Recommendation, 2009. URL: <https://www.w3.org/TR/skos-reference/>.
- [5] ISO, ISO 25964-2:2013 information and documentation—thesauri and interoperability with other vocabularies—part 2, 2013. ISO/TC 46/SC 9, Geneva.
- [6] N. Matentzoglou, J. P. Balhoff, S. M. Bello, et al., A simple standard for sharing ontological mappings (SSSOM), Database 2022 (2022) baac035. doi:10.1093/database/baac035.
- [7] J. Hernandez, D. Golpayegani, D. Lewis, An open knowledge graph-based approach for mapping concepts and requirements between the EU AI Act and international standards, AI and Ethics (2025). doi:10.1007/s43681-025-00708-6.
- [8] OASIS, OSLC requirements management version 2.1, OASIS Committee Specification, 2016. URL: <https://docs.oasis-open.org/oslc-domains/oslc-rm/v2.1/>.
- [9] R. Hoekstra, J. Breuker, M. Di Bello, A. Boer, LKIF core: Principled ontology development for the legal domain, in: Law, Ontologies and the Semantic Web, IOS Press, 2009, pp. 21–52.
- [10] M. Palmirani, F. Vitali, Akoma ntoso for legal documents, in: Legislative XML for the Semantic Web, Springer, 2011, pp. 75–100.
- [11] M. Palmirani, M. Martoni, A. Rossi, C. Bartolini, L. Robaldo, PrOnto: Privacy ontology for legal reasoning, in: EGOVIS 2018, LNCS 11032, Springer, 2018, pp. 139–152.
- [12] H. J. Pandit, B. Esteves, G. P. Krog, et al., Data privacy vocabulary (DPV)—version 2, W3C Community Group Report, 2024. URL: <https://w3id.org/dpv>.
- [13] E. Francesconi, G. Governatori, Patterns for legal compliance checking in a decidable framework of linked open data, Artificial Intelligence and Law (2022). doi:10.1007/s10506-022-09317-8.
- [14] H. J. Pandit, D. O’Sullivan, D. Lewis, Queryable provenance metadata for GDPR compliance, Procedia Computer Science 137 (2018) 262–268. doi:10.1016/j.procs.2018.09.026.
- [15] D. Golpayegani, H. J. Pandit, D. Lewis, AIRO: An ontology for representing AI risks based on the proposed EU AI act and ISO risk management standards, in: Towards a Knowledge-Aware AI, IOS Press, 2022, pp. 51–65.
- [16] M. D. Wilkinson, et al., The FAIR guiding principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. doi:10.1038/sdata.2016.18.
- [17] T. Lebo, et al., PROV-O: The PROV ontology, W3C Recommendation, 2013. URL: <https://www.w3.org/TR/prov-o/>.

A. Online Resources

- AIDGO,