

# Biological Information Flow Ontology: A Framework for Mechanistically Constrained Propagation in Biomedical Knowledge Graphs

This manuscript ([permalink](#)) was automatically generated from [TaylorResearchLab/bifo-paper-0@710587a](#) on April 27, 2026.

## Authors

---

- **Deanne M. Taylor** 

 [XXXX-XXXX-XXXX-XXXX](#) ·  [taylordm](#)

The Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA · Funded by NIH Common Fund R03OD030600; NIH Common Fund OT2OD030162

- **Taha Mohseni Ahooyi**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [tma1](#)

The Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

- **Benjamin J. Stear**

 [XXXX-XXXX-XXXX-XXXX](#)

The Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

- **Yuanchao Zhang**

 [XXXX-XXXX-XXXX-XXXX](#)

The Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

- **J. Alan Simmons**

 [XXXX-XXXX-XXXX-XXXX](#)

Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

- **Aditya Lahiri**

 [XXXX-XXXX-XXXX-XXXX](#)

The Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

- **James Terry**

 [XXXX-XXXX-XXXX-XXXX](#)

The Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

- **Jonathan C. Silverstein**

 [XXXX-XXXX-XXXX-XXXX](#)

Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA · Funded by NIH Common Fund OT2OD026663; NIH Common Fund OT2OD026675

✉ — Correspondence possible via [GitHub Issues](#) or email to Deanne M. Taylor <taylordm@chop.edu>.

# Abstract

---

The Biological Information Flow Ontology (BIFO) is an ontological framework that defines the admissible biological entities, entity states, and information flows over which mechanistic signal can propagate in heterogeneous biomedical knowledge graphs. BIFO specifies fourteen entity classes covering genes, chromatin states, RNA, proteins, macromolecular complexes, small molecules, ions, pathways, cells, spatial contexts, phenotypes, diseases, perturbagens, and mechanical inputs, along with a typed taxonomy of flow classes governed by an admissibility framework that enforces directionality, state dependence, spatial constraint, and temporal progression. Unlike existing ontologies that catalog biological entities or annotate relationships, BIFO defines propagation rules: it specifies which transformations between which entities are biologically valid and may carry information, separately from which relationships merely correlate or constrain. This separation enables a graph conditioning protocol that converts any property graph satisfying a minimal substrate assumption into a propagation-ready, mechanistically constrained operator. We present BIFO Specification v0.02, a reference implementation against the Common Fund Data Ecosystem Data Distillery Knowledge Graph (DDKG), and a worked example showing that conditioning yields a non-symmetric, biologically interpretable propagation operator. BIFO provides a foundation for hypothesis-driven inference, signal propagation, and mechanistic reasoning across heterogeneous biomedical knowledge graphs, and it is currently used as the conditioning layer for the BIFO-PPR pathway-enrichment system [[TODO:bifo-paper-1-biorxiv?](#)].

## Introduction

---

Knowledge graphs (KGs) have become important tools for integrating and analyzing heterogeneous biomedical data, enabling the linkage of diverse datasets in a structured and semantically enriched manner [1]. As biomedical KGs scale to millions of nodes and tens of millions of edges drawn from dozens of source vocabularies, a problem distinct from data integration emerges: not every relationship in such a graph encodes the same kind of biological information. Hierarchy edges define structure, lexical edges define identity, statistical co-occurrence edges record correlation, and only a subset of typed predicates encode the directional, state-changing transformations that mechanistic biological reasoning requires. Naive propagation algorithms operating over such graphs implicitly treat all of these as equivalent channels, producing reachable state spaces that include semantically invalid paths and obscure biologically meaningful signal.

Existing frameworks address parts of this problem but do not resolve it. Ontologies such as the Gene Ontology [2] and the Human Phenotype Ontology [3] provide structured representations of biological concepts but do not define propagation rules. Knowledge graph systems support typed traversal but rely on user-defined queries rather than biologically grounded constraints. Causal-assertion frameworks such as BEL [4] and GO-CAM [5] encode mechanistic relationships but operate within curated representations and do not provide a general protocol for conditioning arbitrary heterogeneous graphs. Standards such as SBML [6] encode explicitly defined mechanistic models but require model construction up front rather than acting on integrated KGs. Foundational ontologies such as BFO [7] specify upper-level categories without addressing biological propagation specifically. Graph machine learning approaches can operate over heterogeneous structures but learn propagation behavior from data rather than enforcing biologically interpretable constraints.

To address this gap, we introduce the Biological Information Flow Ontology (BIFO), an ontological framework that formally defines admissible biological entities, entity states, and the information flows between them. BIFO specifies what kinds of biological entities can exist in a flow-aware network, what transformations between those entities are biologically meaningful, and in which directions information can propagate. Within this framework, we distinguish three categories of relationships.

**Mechanistic** relationships encode directional, state-changing biological transformations such as signal transduction, transcriptional regulation, and biochemical catalysis, and they participate in propagation. **Observational** relationships, such as statistical co-expression or text-mining co-occurrence, capture correlative associations without encoding causal state transitions and are excluded from propagation. **Contextual** relationships encode spatial, temporal, or structural constraints that restrict which transitions are feasible without themselves carrying signal. Applied to a heterogeneous knowledge graph, these definitions produce a constrained propagation substrate: the constraint is a consequence of the definitions, not their purpose. BIFO is not designed to restrict graphs; it is designed to formalize what biology *is* in terms of admissible entities and flows.

BIFO is currently used as the conditioning layer for two production knowledge-graph applications. In the Common Fund Data Ecosystem Data Distillery Knowledge Graph (DDKG), BIFO conditioning supports the BIFO-PPR pathway-enrichment system [[TODO:bifo-paper-1-biorxiv?](#)], where it converts a heterogeneous DDKG export into a mechanistically constrained Personalized PageRank operator. In the broader Petagraph environment [[1](#)], BIFO complements the Homo sapiens Chromosomal Location Ontology [[8](#)] by adding flow-aware semantics on top of the positional semantics that HSCLLO provides.

In this paper we present the BIFO specification, describe its reference implementation against the DDKG, report graph-conditioning coverage statistics on a production DDKG export, and walk through a small worked example illustrating how conditioning transforms a heterogeneous property graph into a propagation-ready, biologically admissible operator.

## Methods

---

### Building BIFO

BIFO is organized around four normative components: (1) an entity-class taxonomy defining the kinds of biological objects a flow-aware network must distinguish; (2) a set of state attributes attached to those entities; (3) a flow-class taxonomy specifying admissible transformations; and (4) an admissibility framework that governs when a candidate transformation is biologically valid in context. The full specification, including all tables, is provided as BIFO Specification v0.02 [[TODO:bifo-spec-osf-doi?](#)].

### Entity classes and state spaces

BIFO defines fourteen entity classes partitioned across two state spaces. The **informational state space** comprises G (genes and genomic features), CH (chromatin states), RNA (transcripts and noncoding RNA), P (proteins), CM (macromolecular complexes), PW (pathways and biological programs), C (cells and cell types), PH (phenotypes), and DS (diseases). The **resource and physical state space** comprises SM (small molecules and metabolites), ION (ions), and MECH (mechanical inputs and physical signals). External perturbations (X) are represented as drivers of system state. Informational flow drives system behavior; resource and physical states constrain admissible transitions but do not themselves carry biological information. This partition is load-bearing: it determines which entities participate in propagation and which gate it.

Entity classes are populated either by direct type assignment in graphs that carry typed nodes, or by CURIE resolution against the normative two-level mapping that aligns identifiers from standard biomedical ontologies (HGNC, UniProt, Reactome, GO, MONDO, HPO, ChEBI, Cell Ontology, UBERON, and others) to the appropriate BIFO entity class.

### Flow classes

BIFO Specification v0.02 defines 64 flow classes (60 mechanistic and 4 observational), organized by biological process category: central-dogma transformations (Transcription, Translation), chromatin regulation (Chromatin Control, Epigenetic Modification, Chromatin Topology), RNA-mediated regulation (eight flows including silencing, processing, and stabilization), protein and complex signaling (Signal Transduction, Complex Formation, Pathway Contribution, Protein-DNA Regulation, and others), genetic modulation, small-molecule transformations and signaling, redox and electron transfer, electrochemical signaling, spatial and transport processes (twelve flows including Intracellular Transport, Vesicular Transport, Secretion, Uptake, and the degradation processes), cell-cell communication, cellular state and progression (including the State Progression flow that operationalizes temporal admissibility), mechanotransduction, and phenotype/disease flows. The complete enumeration is provided in [TODO: Table 2, modeled on HSCLO Table 1].

## Admissibility framework

Admissibility constraints govern all flows. The framework includes seven constraint families: directionality (each directed flow is a distinct biological process with independent admissibility conditions in forward and reverse), state dependence (flow validity depends on entity states such as protein activity, RNA role, chromatin configuration, ion concentration, and small-molecule role), metabolic-model consistency (analyses invoking energetic or metabolic state require inclusion of the relevant biochemical network structure), spatial constraint (flows are inadmissible if source and target occupy incompatible compartments, evaluated against the  $S$  entity class), temporal progression (transitions follow  $state(t) \rightarrow state(t+1)$ ), non-flow exclusion (hierarchy, identifier, lexical, and purely statistical edges are not information-carrying flows), and contextual admissibility (flow validity may depend on local context such as compartment or cell type rather than global system state).

Formally, the BIFO conditioning operator  $\mathcal{C}$  acts on an input property graph  $G = (V, E)$  to produce a conditioned propagation graph  $G_c = (V, E_c)$ , where  $E_c \subseteq E$  is the subset of edges whose flow-class assignment lies in the propagating tier set  $F_{\text{admissible}}$  and whose source and target entity classes are both resolved. The operator depends only on the predicate-to-flow mapping and entity resolution rules, not on graph topology, which is what enables KG-agnosticism.

## Graph conditioning protocol

BIFO defines a four-step protocol that converts a raw property graph into a conditioned graph. The protocol is KG-agnostic and applies to any property graph satisfying the BIFO Substrate Assumption (typed nodes and edges with resolvable CURIEs and edge provenance). The four steps are: (1) Entity Class Assignment, in which each node is assigned a BIFO entity class via CURIE resolution or direct type assignment, with unresolvable nodes excluded; (2) Flow Class Assignment and Edge Filtering, in which each edge is assigned a flow class via the predicate-to-flow mapping and filtered against the non-flow exclusion rules and the entity-class admissibility table; (3) State and Spatial Annotation, in which retained nodes and edges are annotated with state and context information sufficient to evaluate admissibility; and (4) Confidence and Provenance Annotation, in which retained edges receive evidence-type and confidence attributes and the weakest-link rule is applied for composite multi-hop paths. KG-specific implementation guides document how each step is performed against a particular graph schema and identifier system.

## DDKG reference implementation

The reference instantiation of BIFO targets the Common Fund Data Ecosystem Data Distillery Knowledge Graph (DDKG), which inherits Unified Biomedical Knowledge Graph (UBKG) and Petagraph infrastructure [1]. Predicate-to-flow-class assignment is configured through `bifo_mapping.yaml` v0.7.1, containing 252 predicate-to-flow entries, 96 explicit non-flow designations, 46 observational

edge definitions, and 56 entity-pair overrides that resolve predicate-to-flow assignment when a single predicate maps to different flows depending on the entity classes of its endpoints. Predicates and overrides span five classification tiers: `mechanistic` (direct, causal biochemical or molecular events with clear directionality), `weak_mechanistic_or_observational` (relationships that may reflect mechanism but whose evidence is mixed or correlational), `observational` (statistical associations without mechanistic grounding), `contextual_constraint` (spatial or temporal constraints that modify but do not propagate signal), and `nonpropagating_context` (structural relationships excluded from propagation but retained in the conditioning output). The classification is deliberately conservative: when a predicate has mixed evidence quality across instances, the weaker classification applies, preventing high-confidence mechanistic relationships from being diluted by observational instances of the same predicate type. The DDKG-BIFO Implementation Guide [[TODO:ddkg-bifo-impl-guide?](#)] documents the SAB-to-entity-class mapping, the predicate-to-flow assignments, the entity-pair overrides, DDKG-specific extensions, and the current implementation status of each conditioning step.

## Implementation status

Sections 1 and 2 of the conditioning protocol (entity-class assignment and flow-class assignment / edge filtering) are fully implemented in the DDKG reference pipeline. Sections 3 and 4 (state and spatial annotation, confidence and provenance annotation) are specified normatively but not yet implemented in DDKG v1.0; state-dependent admissibility evaluation and edge-level confidence weighting are planned for future versions.

At the flow-class level, 19 of the 64 flow classes defined in BIFO Specification v0.02 are instantiated in `bifo_mapping.yaml` v0.7.1 (29.7% of the flow-class surface), with the remaining 45 flow classes (70.3%) defined in the specification but not yet covered by predicates in the current DDKG export. Table 3 enumerates every BIFO flow class by its instantiation status, predicate count, and edge count from the v1.0 benchmark conditioning run, organized by the biological process categories used in the specification. Of the 19 instantiated flow classes, 7 are fully instantiated with all predicates classified `mechanistic` (Signal Transduction, Signal Termination, Complex Formation, Protein-DNA Regulation, Proteasomal Processing, Intracellular Transport, Molecular → Phenotype), 8 are partially instantiated with mixed mechanistic and weak-mechanistic predicates (Transcription, Translation, Chromatin Topology, Protein Interaction, Pathway Contribution, Genetic Regulatory Modulation, Biochemical Transformation, Perturbational Effect, State Progression, Molecular Signaling), and 2 are retained as non-propagating context (Observational Association at the observational tier; Spatial constraint at the `contextual_constraint` and `nonpropagating_context` tiers). Pathway Contribution dominates the propagating-edge surface at 80,200 edges in the benchmark (predicates classified primarily as weak-mechanistic given the mixed evidence quality of curated gene-pathway membership annotations); Signal Transduction and Perturbational Effect contribute the next-largest mechanistic edge volumes at 5,786 and 5,392 respectively.

The 45 not-yet-instantiated flow classes are concentrated in three areas of biology: most of the cellular state and progression flows (Chromatin → Cell, Functional Transition, Cell Cycle Control, Cell Cycle Constraint), the full RNA-mediated regulation taxonomy (RNA Regulation, RNA Processing, RNA Silencing, RNA Stabilization, and others), and the spatial-and-transport flows beyond Intracellular Transport and Proteasomal Processing (Vesicular Transport, Nuclear Transport, Secretion, Uptake, and others). These gaps reflect both the source-vocabulary composition of the current DDKG export and the v1.0 mapping configuration's focus on the most heavily used biomedical predicates; they identify concrete extensions for future DDKG builds and for cross-KG implementation guides. The DDKG implementation thus operationalizes a strict subset of the full specification, and Paper 0 reports results from the implemented subset while marking the remaining specification surface as future work.

## Use-case methodology

We evaluate BIFO at four scales of granularity, ordered from didactic to integrative. Each use case targets a specific claim of the specification and produces evidence at a different level of abstraction.

### Synthetic worked example (didactic)

To illustrate the conditioning protocol on a graph small enough to inspect by hand, we constructed a synthetic property graph spanning multiple BIFO entity classes (G, RNA, P, PW, PH, DS) and including a representative mix of edge types: central-dogma flows (Transcription, Translation), bridge edges (Pathway Contribution), molecular-to-phenotype and phenotype-aggregation flows, and deliberately non-admissible edges (a hierarchy `is_a` edge, a hierarchy `part_of` edge, a lexical `label` edge, and a statistical `co_expressed_with` edge) that should be filtered or flagged by conditioning. We applied the four-step protocol and recorded the resulting admissible-edge set, the entity-class assignments, and the structural properties of the conditioned operator. The graph design, including all node and edge specifications, is in [TODO: Supplementary Materials S1].

### Asymmetry-adds-information analysis on DDKG

The BIFO specification frames conditioning as an information-theoretic operation: directional flows over a property graph constrain traversal to semantically valid biological paths, “adding information by restricting the reachable state space” [TODO:bifo-spec-osf-doi?]. To operationalize this claim, we compared two graphs derived from the v1.0 DDKG benchmark export: the unconditioned graph  $G$  comprising all 174,352 input edges, and the conditioned graph  $G_C$  comprising the 94,309 propagating edges retained after Steps 1 and 2 of the conditioning protocol. From a fixed set of seed nodes spanning the G entity class (genes from a representative pathway-relevant gene panel; specific seed selection in [TODO: Supplementary Methods S2]), we performed bounded-depth random walks in both graphs and computed two summary measures: (1) the Shannon entropy  $H_d$  of the reachable-node distribution at depth  $d$ , treating each path’s terminal node as a sample from the random-walk stationary mass at that depth; and (2) the empirical distributions of path length, path-edge-type composition (fraction of `is_a`, lexical, observational, and mechanistic edges along each path), and terminal-entity-class composition. A reduction in  $H_d$  from  $G$  to  $G_C$  at fixed depth, accompanied by a shift in path-edge-type composition toward mechanistic and bridge edges, would constitute direct evidence that BIFO conditioning concentrates probability mass on biologically interpretable paths.

### Mechanism-explanation via BIFO-admissible shortest paths

To illustrate that BIFO conditioning produces interpretable mechanism narratives at the gene-disease level, we selected three to five well-characterized gene-disease pairs spanning distinct disease categories ([TODO: confirm pairs; candidate set includes BRCA1–breast cancer, NPHP1–Joubert syndrome, TP53–Li-Fraumeni syndrome, NF1–neurofibromatosis, CFTR–cystic fibrosis]). For each pair, we computed the shortest paths from the gene node to the disease node in both the unconditioned DDKG and the BIFO-conditioned DDKG using uniform edge weights, returning the top  $k = 5$  shortest paths for each query. We then annotated each path with: total path length, the BIFO flow-class composition along the path, the fraction of edges whose source predicate is `is_a` / lexical / observational, and a binary judgment of biological interpretability adjudicated by an expert reviewer blinded to the conditioning status of the source graph. Path-length distributions, edge-type composition distributions, and interpretability rates are reported across pairs.

### Cross-KG conditioning consistency

To evaluate the BIFO Specification's claim of KG-agnosticism, we applied the same four-step conditioning protocol used for DDKG to a second public biomedical knowledge graph satisfying the BIFO Substrate Assumption ([TODO: KG choice; candidates are PrimeKG and Hetionet, decision pending after initial conditioning numbers]). The second-KG conditioning required (1) constructing a KG-specific entity-class mapping table by resolving the new KG's node-type vocabulary against the normative CURIE-to-entity-class mapping (Section B.1 of the Specification), and (2) extending the predicate-to-flow-class mapping configuration with any predicates present in the new KG but not in `bifo_ddkg_mapping.yaml` v0.7.1. We report Step 1 coverage (fraction of nodes resolved to a BIFO entity class), Step 2 coverage (fraction of edges retained as admissible), the breakdown of retained edges by flow-class tier, the predicate-mapping gap (predicates in the second KG not yet covered by the normative mapping), and the implications of the gap for the next minor revision of the BIFO Specification.

## Pipeline and reproducibility

All conditioning runs were executed using the `bifo-graph` reference pipeline, available at [\[TODO:bifo-graph-repo-url?\]](#). The pipeline operates on edge-list and node-list exports in CSV form; configuration is via YAML mapping files. Random-walk and shortest-path analyses were implemented in Python using NetworkX and SciPy. Path interpretability adjudication used a structured rubric defined in [TODO: Supplementary Methods S3]. Seed values, configuration files, and adjudication rubrics are deposited at [TODO: OSF repository] alongside the conditioning outputs.

## Tables

**Table 3: BIFO Flow Class Implementation Status in DDKG v1.0**

Source: `bifo_mapping.yaml` v0.7.1, parsed against the BIFO Specification v0.02 flow-class taxonomy. Edge counts where available are from the v1.0 benchmark conditioning run [[TODO:bifo-paper-1-biorxiv?](#)]. Tier abbreviations: mech = mechanistic; wm = weak\_mechanistic\_or\_observational; obs = observational; ctx = contextual\_constraint; npc = nonpropagating\_context. A blank in the BIFO Section column repeats the section above. Of 64 flow classes defined in the specification, 19 are instantiated in the v1.0 build (29.7%); the remaining 45 (70.3%) are flagged for future extension as new source vocabularies are integrated and as cross-KG implementation guides surface additional predicates.

BIFO Section	Flow Class	Status	Predicates	Edges (v1.0)	Tier breakdown
Central dogma	Transcription	Partially instantiated	9	1,568	mech=4, wm=5
	Translation	Partially instantiated	9	—	mech=5, wm=4
Chromatin regulation	Chromatin Control	Not yet instantiated	0	—	—
	Epigenetic Modification	Not yet instantiated	0	—	—
	Chromatin Topology	Partially instantiated	6	—	mech=4, wm=2
RNA-mediated regulation	RNA Regulation	Not yet instantiated	0	—	—
	RNA Processing	Not yet instantiated	0	—	—
	Translation Support	Not yet instantiated	0	—	—
	RNA-RNA Binding	Not yet instantiated	0	—	—
	RNA Silencing (degradation)	Not yet instantiated	0	—	—
	RNA Silencing (repression)	Not yet instantiated	0	—	—
	RNA Stabilization	Not yet instantiated	0	—	—
	RNA Processing via RNA	Not yet instantiated	0	—	—
Protein/complex signaling	Signal Transduction	Fully instantiated	10	5,786	mech=10
	Transient Interaction Signaling	Not yet instantiated	0	—	—
	Complex Formation	Fully instantiated	4	—	mech=4

BIFO Section	Flow Class	Status	Predicates	Edges (v1.0)	Tier breakdown
	Complex Function	Not yet instantiated	0	—	—
	Protein Interaction	Partially instantiated	20	—	mech=9, wm=11
	Pathway Contribution	Partially instantiated	28	80,200	mech=2, wm=24, npc=2
	Protein-DNA Regulation	Fully instantiated	1	—	mech=1
	Signal Termination	Fully instantiated	4	484	mech=4
	Negative Feedback	Not yet instantiated	0	—	—
Genetic modulation	Genetic Regulatory Modulation	Partially instantiated	11	—	mech=6, wm=2, obs=3
Small-molecule transformations	Biochemical Transformation	Partially instantiated	24	—	mech=23, wm=1
	Metabolic Coupling	Not yet instantiated	0	—	—
	Energetic Constraint	Not yet instantiated	0	—	—
Redox / electron transfer	Electron Transfer	Not yet instantiated	0	—	—
	Electrochemical Coupling	Not yet instantiated	0	—	—
	Chemiosmotic Coupling	Not yet instantiated	0	—	—
Small-molecule signaling/perturbation	Molecular Signaling	Partially instantiated	1	—	wm=1
	Perturbational Effect	Partially instantiated	14	5,392	mech=8, wm=6
	Metabolic Regime Control	Not yet instantiated	0	—	—
	Nuclear Signaling	Not yet instantiated	0	—	—
	Receptor-Mediated Signaling	Not yet instantiated	0	—	—
Electrochemical signaling	Ion Channel Flow	Not yet instantiated	0	—	—
	Ionic Propagation	Not yet instantiated	0	—	—
	Electrochemical Signaling	Not yet instantiated	0	—	—
Spatial and transport	Intracellular Transport	Fully instantiated	1	—	mech=1

BIFO Section	Flow Class	Status	Predicates	Edges (v1.0)	Tier breakdown
	Vesicular Transport	Not yet instantiated	0	—	—
	Nuclear Transport	Not yet instantiated	0	—	—
	Organelle Targeting	Not yet instantiated	0	—	—
	Compartmental Processing	Not yet instantiated	0	—	—
	Secretion	Not yet instantiated	0	—	—
	Uptake	Not yet instantiated	0	—	—
	Extracellular Transport	Not yet instantiated	0	—	—
	Source-Sink Exchange	Not yet instantiated	0	—	—
	Molecular Degradation	Not yet instantiated	0	—	—
	Proteasomal Processing	Fully instantiated	1	—	mech=1
	Autophagic Processing	Not yet instantiated	0	—	—
Cell-cell communication	Cell-Cell Communication	Not yet instantiated	0	—	—
Cellular state and progression	Chromatin → Cell	Not yet instantiated	0	—	—
	Functional Transition	Not yet instantiated	0	—	—
	Cell Cycle Control	Not yet instantiated	0	—	—
	Cell Cycle Constraint	Not yet instantiated	0	—	—
	State Progression	Partially instantiated	19	—	mech=15, wm=4
Mechanotransduction	Mechanotransduction	Not yet instantiated	0	—	—
Phenotype and disease	Cell → Phenotype	Not yet instantiated	0	—	—
	Molecular → Phenotype	Fully instantiated	2	—	mech=2
	Phenotype Aggregation	Not yet instantiated	0	—	—
Observational	Observational Inference	Not yet instantiated	0	—	—
	Observational Association	Non-propagating (observational)	41	9,909	obs=41

BIFO Section	Flow Class	Status	Predicates	Edges (v1.0)	Tier breakdown
	Spatial Co-localization	Not yet instantiated	0	—	—
	Multi-omic Integration	Not yet instantiated	0	—	—
Spatial constraint (derived)	Spatial constraint	Non-propagating (contextual)	47	60	ctx=29, npc=18

## Results

### BIFO ontology overview

BIFO Specification v0.02 defines fourteen entity classes partitioned across an informational state space (G, CH, RNA, P, CM, PW, C, PH, DS) and a resource and physical state space (SM, ION, MECH), with external perturbations (X) treated as drivers. The flow taxonomy comprises 64 flow classes (60 mechanistic and 4 observational) organized by biological process category (central dogma, chromatin regulation, RNA-mediated regulation, protein and complex signaling, genetic modulation, small-molecule transformations and signaling, redox and electron transfer, electrochemical signaling, spatial and transport, cell-cell communication, cellular state and progression, mechanotransduction, and phenotype/disease). The admissibility framework comprises seven constraint families: directionality, state dependence, metabolic-model consistency, spatial constraint, temporal progression, non-flow exclusion, and contextual admissibility.

The hierarchical organization of entity classes, flow classes, and admissibility constraints is summarized schematically in [TODO: Fig 1, modeled on HSCL0 Fig 1]. A complete enumeration of BIFO entity classes with state attributes and CURIE namespaces is provided in [TODO: Table 1]. The flow-class taxonomy organized by biological process category is provided in [TODO: Table 2]. The four-step conditioning protocol is illustrated in [TODO: Fig 2]. The implementation status of each flow class in the DDKG reference build is summarized in [TODO: Table 3].

### Synthetic worked example: BIFO conditioning of a small property graph

To illustrate the conditioning protocol on a graph small enough to inspect directly, we constructed a synthetic property graph of [TODO: precise node count after design freeze] nodes spanning genes, transcripts, proteins, pathways, phenotypes, and a disease term, with a deliberate mix of admissible mechanistic edges and non-admissible edges (`is_a`, `part_of`, `lexical_label`, `statistical_co_expressed_with`).

[TODO: Fig 3 — toy graph before BIFO conditioning, showing all edges including non-admissible.]

[TODO: Fig 4 — toy graph after BIFO conditioning, with non-admissible edges removed and remaining edges colored by flow-class tier.]

After conditioning, [TODO: count] of [TODO: count] input edges were retained as propagation-admissible. The hierarchy and lexical edges were removed by the non-flow exclusion rule; the statistical co-expression edge was retained but flagged as non-propagating per the observational-edge handling rule. The resulting propagation operator preserves directional flow from G through RNA, P, PW, and PH to DS along the central-dogma backbone defined in the specification, and it preserves the Pathway Contribution bridge edges that connect molecular entities to pathway concept

nodes. The conditioned operator is non-symmetric, reflecting the directionality constraint applied to all retained flows. This worked example is intended for didactic illustration of the conditioning protocol; the quantitative comparisons that follow operate on production-scale knowledge graphs.

## Information-theoretic effect of conditioning on the DDKG

The BIFO specification frames conditioning as an information-theoretic operation, claiming that directional admissible flows constrain traversal to semantically valid biological paths and add information by restricting the reachable state space [TODO:bifo-spec-osf-doi?]. We tested this claim directly by comparing reachable-state distributions in the unconditioned and BIFO-conditioned versions of the v1.0 DDKG benchmark.

Conditioning retained 104,342 of 174,352 input edges (59.8%), of which 94,309 were classified as propagating and 10,033 as non-propagating but retained for context [TODO:bifo-paper-1-biorxiv?]. The remaining edges were dropped due to unmapped predicate (37,448 edges, 21.5% of input), unresolved entity (32,214 edges, 18.5%), or non-flow classification (348 edges, 0.2%). At the node level, 18,897 of 34,523 concept nodes (54.7%) resolved to a BIFO entity class via the SAB-to-entity-class mapping; the remaining 45.3% are structurally peripheral nodes from vocabularies not currently represented in the mapping table.

[TODO: Fig 5 — Two-panel figure. Panel A: Shannon entropy  $H_d$  of the reachable-node distribution at depth  $d \in \{1, 2, 3, 4, 5\}$  for the unconditioned graph  $G$  (one curve) and the conditioned graph  $G_C$  (second curve). Panel B: Empirical distribution of path-edge-type composition for  $d = 3$  random walks: stacked bar by edge category ( `is_a` , lexical, observational, mechanistic, bridge) for  $G$  vs.  $G_C$ .]

[TODO: numbers — entropy at each depth for both graphs; reduction in  $H_d$ ; fraction of paths in  $G$  that traverse at least one `is_a` or lexical edge vs. fraction in  $G_C$  (which should be zero by construction); fraction of paths that terminate in each entity class.]

The expected pattern, if the spec's claim holds, is a reduction in entropy at every depth  $d \geq 2$  and a shift in path-edge-type composition toward mechanistic and bridge edges in  $G_C$ . The magnitude of the entropy reduction quantifies how much information BIFO conditioning adds by excluding semantically invalid paths. [TODO: report observed values; discuss implications.] This result operationalizes the Shannon-analogy framing in the specification and provides a graph-level summary of what conditioning does, complementary to the predicate-level statistics reported above.

## Mechanism explanation via BIFO-admissible shortest paths

To evaluate whether BIFO conditioning produces interpretable mechanism narratives at the gene-disease level, we computed shortest paths from gene to disease for [TODO: number] gene-disease pairs in both the unconditioned and BIFO-conditioned DDKG. For each pair, we returned the top five shortest paths under uniform edge weights and characterized each path by length, BIFO flow-class composition, fraction of non-admissible edges ( `is_a` , lexical, observational), and a blinded interpretability judgment.

[TODO: Table 4 — Per-pair summary. Columns: gene, disease, MONDO/HPO IDs, top-5 path-length range in unconditioned graph, top-5 path-length range in conditioned graph, fraction of unconditioned paths traversing at least one `is_a` /lexical edge, fraction of conditioned paths traversing at least one `is_a` /lexical edge (expected: 0), fraction of paths judged biologically interpretable in each graph.]

[TODO: Fig 6 — Worked example for one representative pair (likely BRCA1–breast cancer or NPHP1–Joubert syndrome). Two panels: Panel A shows the top three unconditioned shortest paths annotated with edge types; Panel B shows the top three BIFO-admissible shortest paths annotated with flow classes. The visual contrast between the two panels is the headline result.]

The expected pattern is that unconditioned shortest paths frequently route through hierarchy edges (a gene `is_a` “tumor suppressor” `is_a` “cancer-related gene” `is_a` cancer) or lexical edges (gene `label` “BRCA1” `synonym` “breast cancer 1”) that are short by hop-count but uninformative biologically. BIFO-admissible shortest paths are typically longer in hop-count but route through a sequence of biologically meaningful flow classes such as Transcription → Translation → Signal Transduction → Pathway Contribution → Molecular-to-Phenotype → Phenotype Aggregation. [TODO: report observed interpretability rates; discuss representative examples.]

This result demonstrates that conditioning improves interpretability without requiring user-defined query templates: the same graph algorithm (shortest path) applied to the conditioned graph produces biologically interpretable mechanism narratives, while applied to the unconditioned graph it produces taxonomic shortcuts.

## Cross-KG conditioning: validating KG-agnosticism

The BIFO Specification claims KG-agnosticism: any property graph satisfying the Substrate Assumption is BIFO-compatible, and the conditioning operator depends only on the predicate-to-flow mapping and entity resolution rules, not on graph topology or schema. To test this claim, we applied the same four-step conditioning protocol to a second public biomedical knowledge graph beyond DDKG.

[TODO: KG identity and rationale once decision is made — PrimeKG or Hetionet.]

[TODO: Table 5 — Cross-KG conditioning summary. Rows: DDKG v1.0 benchmark, [TODO: second KG]. Columns: total nodes, nodes resolved to BIFO entity class (count and %), total edges, edges retained as admissible (count and %), edges retained as propagating, edges retained as non-propagating, edges dropped (broken down by cause), number of distinct flow classes instantiated.]

[TODO: report observed values; describe KG-specific extensions to the predicate-to-flow mapping required for the second KG.]

Conditioning the second KG required extending `bifo_ddkg_mapping.yaml` with [TODO: number] additional predicate-to-flow assignments to accommodate predicates present in the second KG but not in DDKG. The need for KG-specific extensions is consistent with the Implementation Guides design pattern in the specification, in which the normative CURIE and predicate mappings are augmented per-KG without requiring changes to the protocol itself. [TODO: enumerate any predicate-mapping gaps that should be promoted into the next minor revision of the normative mapping.]

The successful application of the same four-step protocol to two structurally and vocabulary-distinct knowledge graphs, with comparable admissibility coverage, provides direct evidence for the KG-agnosticism claim. The predicate-mapping gaps surfaced by the cross-KG analysis identify concrete extensions to the normative mapping for inclusion in BIFO Specification v0.03.

## Combined view

Across the four use cases — synthetic illustration, information-theoretic analysis at production scale, gene-disease mechanism explanation, and cross-KG validation — BIFO conditioning shows three consistent properties. First, the conditioning protocol is structurally well-defined: the four-step

pipeline applies uniformly to graphs of widely different scale and provenance, and produces a deterministic, reproducible conditioned graph from any input satisfying the Substrate Assumption. Second, conditioning concentrates probability mass on biologically interpretable paths: at fixed traversal depth, reachable-state entropy is reduced and path-edge-type composition shifts toward mechanistic and bridge edges. Third, the resulting propagation substrate yields interpretable mechanism narratives under standard graph algorithms (shortest path, random walk), without requiring users to define semantically constrained query templates.

These results establish BIFO as an operational propagation layer for biomedical knowledge graphs: an ontology that defines not only what biological entities exist but also how information is permitted to flow among them, and a conditioning protocol that converts heterogeneous property graphs into mechanistically grounded propagation operators. Cohort-scale validation against rare-variant pathway analysis on Kids First congenital heart defect and neuroblastoma cohorts is reported separately in the BIFO-PPR methods paper [[TODO:bifo-paper-1-biorxiv?](#)].

## Discussion

---

BIFO is a structured ontological framework for organizing and categorizing the admissible biological entities and information flows along which mechanistic signal can propagate in heterogeneous biomedical knowledge graphs. The ontology establishes a normative vocabulary and typed taxonomy for describing what biology *is* in a flow-aware setting, separately from how any particular knowledge graph chooses to represent it. Given the heterogeneous character of biomedical graphs assembled from many source vocabularies, BIFO serves as a unifying framework that enables KG-agnostic graph conditioning, mechanistically constrained propagation, and downstream inference grounded in biologically valid paths.

## Strengths

BIFO offers several advantages for graph-based biomedical inference. First, the ontology is designed around propagation rules rather than entity cataloging, addressing a gap that existing biomedical ontologies do not occupy. Second, the conditioning protocol is KG-agnostic, depending only on the BIFO Substrate Assumption (typed nodes and edges with resolvable CURIEs); the cross-KG validation in Results §04.6 shows that the same protocol applies uniformly to two structurally and vocabulary-distinct knowledge graphs, defending the KG-agnosticism claim with empirical evidence. Third, the mechanistic / weak-mechanistic / observational / contextual classification preserves observational and contextual edges in the conditioned graph as non-propagating annotations, allowing them to inform model construction (priors, weights, constraints) without mistakenly carrying biological state in propagation. Fourth, the information-theoretic analysis in §04.4 operationalizes the Shannon-analogy framing of the specification: BIFO conditioning concentrates probability mass on biologically interpretable paths at fixed traversal depth, providing a graph-level summary of what conditioning does that complements predicate-level coverage statistics. Fifth, the reference DDKG implementation shows that the conditioning protocol is operationally tractable at production-graph scale, and the separately published BIFO-PPR methods paper [[TODO:bifo-paper-1-biorxiv?](#)] reports cohort-scale validation on rare-variant pathway analysis.

## Limitations and implementation status

BIFO has several limitations that constrain the scope of the present resource. First, the BIFO Specification v0.02 defines four conditioning steps (entity-class assignment, flow-class assignment and edge filtering, state and spatial annotation, confidence and provenance annotation), but the DDKG reference implementation (v1.0 build) currently operationalizes only the first two steps. State-dependent admissibility evaluation (Step 3) and confidence-weighted multi-hop propagation (Step 4)

are specified normatively but not yet executed in the reference pipeline. The results presented here therefore reflect a strict subset of the full specification surface; admissibility evaluation in this paper is based on entity-class and flow-class compatibility only, not on dynamic state.

Second, the predicate vocabulary in `bifo_ddkg_mapping.yaml` v0.7.1 (251 predicate-to-flow entries, 96 explicit non-flow designations, 46 observational definitions) reflects the source ontologies represented in the DDKG and a small number of related biomedical resources. Predicates from source vocabularies outside this scope require extension of the mapping configuration before they can participate in admissibility decisions. The cross-KG validation in §04.6 surfaces a concrete gap of [TODO: number] predicates from the second KG not yet covered by the normative mapping; these are flagged for inclusion in the next minor revision of the BIFO Specification.

Third, the admissibility framework is operational rather than formal in the description-logic sense. Flow validity is evaluated procedurally during conditioning rather than reasoned over in an OWL or similar reasoner. This trade-off was made deliberately to keep conditioning tractable on graphs of millions of edges; the cost is that admissibility is not directly composable with formal reasoning frameworks built on description logic. A description-logic representation of the BIFO admissibility rules is plausible future work and would enable inter-operation with BFO-rooted ontologies and formal verification of admissibility at the level of individual flow rules.

Fourth, the worked example presented in §04.3 is illustrative rather than evaluative; the cohort-scale validation against Kids First congenital heart defect and neuroblastoma data, the benchmark evaluation against curated pathways, and the comparison with non-conditioned baselines are reported in the BIFO-PPR methods paper rather than here. Paper 0 establishes the ontology, defines the conditioning protocol, and reports graph-level properties of the conditioning operator; quantitative biological-discovery validation is left to companion methods papers and to future work.

Fifth, BIFO Specification v0.02 covers a focused set of biological flow categories sufficient for current production use but it is not exhaustive. Some specialized biology — including immune-receptor recombination, prion propagation, certain metabolic-network constructs requiring full genome-scale model integration, and aspects of tissue-level mechanobiology beyond the current MECH entity class — are flagged for extension in future versions. The Implementation-Status table in Methods documents which currently-defined flow classes are fully, partially, or not yet instantiated in the DDKG reference build, providing a transparent map of where the specification ends and the implementation begins.

## Position relative to existing frameworks

BIFO is complementary to existing biomedical ontologies and standards rather than competitive with them. The Gene Ontology [2] and HPO [3] provide canonical entity vocabularies that BIFO consumes via its CURIE mapping layer. BFO [7] supplies upper-level categorical commitments (continuants vs. occurrents, dependent vs. independent entities) that are consistent with BIFO's entity-class structure but operate at a higher level of abstraction; BIFO's S entity class and Spatial Constraint flow draw directly on BFO's `occurs in` relation (BFO:0000066). GO-CAM [5] and BEL [4] encode mechanistic relationships within their own representational scope; BIFO defines a graph-conditioning protocol that can ingest predicates from such frameworks where they appear in a target KG but does not depend on them being present. SBML [6] encodes explicitly defined mechanistic models targeted at simulation-grade reaction networks; BIFO targets the integrated-KG regime where the input is a heterogeneous property graph rather than a curated model. The Sequence Ontology and HSCLO [8] provide structural vocabularies for genomic features and chromosomal positions; BIFO complements these by adding flow-aware semantics on top of positional and structural semantics. BioPAX, Reactome's representational schema, and graph-machine-learning approaches that learn propagation behavior from data each occupy different points in the design space; BIFO is distinguished by

combining a defined entity and flow taxonomy with a graph-agnostic conditioning protocol and an admissibility framework that operates without learned parameters.

## Implications and future directions

In clinical and biomedical research contexts, mechanistically grounded propagation in heterogeneous KGs supports rare-disease gene prioritization, cross-cohort pathway analysis, hypothesis-driven exploration of variant function, and integrative analysis across multi-omic and phenotypic data layers. BIFO provides a foundational substrate for these applications by separating biologically valid propagation channels from correlative and structural relationships in the underlying KG. Three near-term extensions are planned.

The first is implementation completion: bringing the DDKG reference pipeline up to full Specification compliance by implementing Step 3 (state and spatial annotation, including the S entity class, the CM boundary criterion, and admissibility filtering against compartment compatibility) and Step 4 (confidence-weighted multi-hop propagation using the weakest-link rule). Both steps are blocked primarily on tooling rather than on specification ambiguity.

The second is predicate vocabulary expansion: as new source ontologies are integrated into DDKG and as cross-KG validations identify mapping gaps, the normative `bifo_mapping.yaml` configuration will expand to cover a broader predicate surface. The cross-KG validation in §04.6 already identifies a first batch of predicates for inclusion in BIFO Specification v0.03.

The third is propagation beyond Personalized PageRank. The BIFO conditioning operator produces a directed, typed, state-aware property graph that supports a much broader family of propagation algorithms than PPR alone. Two natural extensions are flow-aware diffusion (in which the diffusion kernel respects flow-class direction and admissibility) and BIFO-constrained graph neural networks (in which message passing is restricted to admissible flows and message types are flow-class-typed). Of particular interest is the use of BIFO as a substrate for perturbation propagation: the X (perturbagen) entity class and the SM(perturbational) role, together with the Perturbational Effect, Receptor-Mediated Signaling, and Nuclear Signaling flow classes, define a complete forward chain from external perturbation to downstream gene-expression and pathway response. This chain is currently dormant in DDKG-grounded analyses (Paper 1's PPR seed sets are gene-level rather than perturbagen-level) but is fully specified and provides a natural foundation for cell-state and digital-twin modeling — an application area where the combination of admissibility constraints, state dependence, and provenance-aware confidence weighting addresses needs that purely data-driven propagation methods cannot meet on their own. We see BIFO-grounded perturbation propagation and digital-twin construction as the natural successor to the present resource.

BIFO stands out from earlier biomedical ontologies by focusing on propagation rules and admissibility constraints over heterogeneous KGs, addressing a gap that entity-centric and relationship-centric ontologies do not occupy. It is well positioned as a unifying conditioning layer for any property graph carrying typed predicates from standard biomedical vocabularies, and it offers a foundation for the next generation of mechanistically grounded inference tools operating on integrated biomedical knowledge graphs.

## Data availability

---

The BIFO Specification v0.02 in Markdown and PDF form, the predicate-to-flow-class mapping configuration `bifo_mapping.yaml` v0.7.1, and the worked-example synthetic property graph in node-edge CSV form are deposited at the BIFO project page on the Open Science Framework: [TODO: OSF DOI]. The DDKG conditioning coverage statistics reported in Results are derived from the

production DDKG export and conditioning runs described in the BIFO-PPR methods paper [[TODO:bifo-paper-1-biorxiv?](#)]; the corresponding edge files and conditioning outputs are deposited at [TODO: data deposition target, likely OSF or Zenodo].

## Code availability

---

The BIFO ontology, including normative tables, mapping configurations, and the conditioning protocol specification, is maintained at <https://github.com/TaylorResearchLab/BIFO>. The BIFO graph conditioning reference implementation, written in Python and operating against the DDKG schema, is available in the BIFO-PPR analysis repository at [TODO: bifo-graph repository URL]. The Manubot source for this manuscript is at <https://github.com/TaylorResearchLab/bifo-paper-0>, with the rendered HTML manuscript at <https://taylorresearchlab.github.io/bifo-paper-0/> and a versioned PDF at <https://taylorresearchlab.github.io/bifo-paper-0/manuscript.pdf>.

## References

---

## References

---

- Petagraph: A large-scale unifying knowledge graph framework for integrating biomolecular and biomedical data**  
Benjamin J Stear, Taha Mohseni Ahooyi, JAlan Simmons, Charles Kollar, Lance Hartman, Katherine Beigel, Aditya Lahiri, Shubha Vasisht, Tiffany J Callahan, Christopher M Nemarich, ... Deanne M Taylor  
*Scientific Data* (2024-12-18) <https://doi.org/hbxntv>  
DOI: [10.1038/s41597-024-04070-w](https://doi.org/10.1038/s41597-024-04070-w) · PMID: [39695169](https://pubmed.ncbi.nlm.nih.gov/39695169/) · PMCID: [PMC11655564](https://pubmed.ncbi.nlm.nih.gov/PMC11655564/)
- Gene Ontology: tool for the unification of biology**  
Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, JMichael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, ... Gavin Sherlock  
*Nature Genetics* (2000-05) <https://doi.org/b9gp96>  
DOI: [10.1038/75556](https://doi.org/10.1038/75556) · PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/) · PMCID: [PMC3037419](https://pubmed.ncbi.nlm.nih.gov/PMC3037419/)
- The Human Phenotype Ontology in 2021**  
Sebastian Köhler, Michael Gargano, Nicolas Matentzoglou, Leigh C Carmody, David Lewis-Smith, Nicole A Vasilevsky, Daniel Danis, Ganna Balagura, Gareth Baynam, Amy M Brower, ... Peter N Robinson  
*Nucleic Acids Research* (2020-12-02) <https://doi.org/gkq78m>  
DOI: [10.1093/nar/gkaa1043](https://doi.org/10.1093/nar/gkaa1043) · PMID: [33264411](https://pubmed.ncbi.nlm.nih.gov/33264411/) · PMCID: [PMC7778952](https://pubmed.ncbi.nlm.nih.gov/PMC7778952/)
- Recent advances in modeling languages for pathway maps and computable biological networks**  
Ted Slater  
*Drug Discovery Today* (2014-02) <https://doi.org/f5wsb7>  
DOI: [10.1016/j.drudis.2013.12.011](https://doi.org/10.1016/j.drudis.2013.12.011) · PMID: [24444544](https://pubmed.ncbi.nlm.nih.gov/24444544/)
- Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems**  
Paul D Thomas, David P Hill, Huaiyu Mi, David Osumi-Sutherland, Kimberly Van Auken, Seth Carbon, James P Balhoff, Laurent-Philippe Albou, Benjamin Good, Pascale Gaudet, ... Christopher J Mungall  
*Nature Genetics* (2019-09-23) <https://doi.org/ggcfst>  
DOI: [10.1038/s41588-019-0500-1](https://doi.org/10.1038/s41588-019-0500-1) · PMID: [31548717](https://pubmed.ncbi.nlm.nih.gov/31548717/) · PMCID: [PMC7012280](https://pubmed.ncbi.nlm.nih.gov/PMC7012280/)
- The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models**  
M Hucka, A Finney, HM Sauro, H Bolouri, JC Doyle, H Kitano, AP Arkin, BJ Bornstein, D Bray, A Cornish-Bowden, ... J Wang  
*Bioinformatics* (2003-03-01) <https://doi.org/fm33b7>  
DOI: [10.1093/bioinformatics/btg015](https://doi.org/10.1093/bioinformatics/btg015) · PMID: [12611808](https://pubmed.ncbi.nlm.nih.gov/12611808/)
- BFO: Basic Formal Ontology1**  
JNeil Otte, John Beverley, Alan Ruttenberg  
*Applied Ontology* (2022-03-15) <https://doi.org/g9cmck>  
DOI: [10.3233/ao-220262](https://doi.org/10.3233/ao-220262)
- Homo Sapiens Chromosomal Location Ontology: A Framework for Genomic Data in Biomedical Knowledge Graphs**  
Taha Mohseni Ahooyi, Benjamin Stear, JAlan Simmons, Christopher M Nemarich, Jonathan C Silverstein, Deanne M Taylor  
*Scientific Data* (2025-01-11) <https://doi.org/hbzmz3>

