Manuscript Title

This manuscript (permalink) was automatically generated from lubianat/quali_phd@17651a3 on November 30, 2021.

Authors

- **John Doe**
  XXXX-XXXX-XXXX-XXXX · johndoe · johndoe
  Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Jane Roe**
  XXXX-XXXX-XXXX-XXXX · janeroe
  Department of Something, University of Whatever; Department of Whatever, University of Something

Abstract

The Human Cell Atlas (HCA) is an international effort aiming at characterizing every cell type of the human body. By the virtue of tecniques such as single-cell RNA sequencing, mass cytometry, and multiplexed in situ hybridization, HCA members are producing cell-level data from virtually all human tissues. This wealth of data can have a significant impact on biomedical research, but only if its content is genuinely interoperable. While ontologies and semantic technologies have emerged as key players in the data interoperability ecosystem, there are still gaps to cover between the technical possibilities and the practical applications in biomedical research. Wikidata is a knowledge graph database emerging as a FAIR (Findable, Accessible, Interoperable and Reusable) repository for biological knowledge. The formatting and deployment of information from the Human Cell Atlas to Wikidata can increase information availability and impact, by inserting the findings in a network containing multiple associations of concepts of all areas of knowledge (within and outside science). Conceptually defining cell types in a general and applicable concept, formalized into a database-compatible format, is a massive theoretical challenge. This PhD project aims at studying our current understanding of cell types for development a comprehensive ontological model in Wikidata for cell types. We will review the single-cell literature, refining and formalizing concepts for cell type delimitation. Furthermore, we will use Natural Language Processing and Machine Learning tools to automate knowledge extraction from scientific articles in the scope of the Human Cell Atlas. In an advanced step, we will apply concepts of network theory to develop tools for user-friendly querying of the database, making the knowledge ready for the academic community.

Introduction

Introduction

The Human Cell Atlas (HCA) Project

The advent of single-cell technologies has ignited the desire of a deep knowledge on cells, the building blocks of life [1]. The Human Cell Atlas (HCA) project, has been a major player in the cell knowledge ecosystem, running since 2017 towards the task to characterize every cell type in the human body [2]. The HCA consortium gathers people from all over the world to tackle different parts of the project, so to have a diverse and equitable account of the cell type diversity. [3]

Building a full atlas of human cells comes with multiple challenges. The project includes the detection, in single cells, of RNA species (scRNA-Seq), chromatin accessibility (scATAC-Seq), and protein markers (primarily by CYTOF), as well as spatial information on cells with multiplexed *in situ* hybridization (such as MERFISH) and imaging mass cytometry [2,4]. Every lab inside the project will contribute with its expertise, providing samples that are representative of human diversity.

HCA is set to revolutionize the biomedical sciences, by creating tools and standards for basic research, as well as allowing better characterization of disease, and thus, ultimately, improving diagnostics and therapy. Its products (data, information, knowledge and wisdom) need to be FAIR: findable, accessible, interoperable and reusable. Data stewardship and data management are growing as core dhttps://www.wikidata.org/wiki/Help:Multilingualemands of the scientific community, ranging from data management plans [5] to specialized personnel [5].

The Human Cell Atlas has a dedicated team for organizing data: the Data Coordination Platform (DCP) [6] [4]. The DCP is responsible for tracing the plan for computational interoperability, from the data generators to the consumers.[4]. The Human Cell Atlas has its portal for data (https://dahttps://www.wikidata.org/wiki/Help:Multilingualta.humancellatlas.org/) which composes the data repository landscape with other resources, like the Broad Institute Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) and the Chan-Zuckerberg Biohub Tabula Sapiens (https://tabula-sapiens-portal.ds.czbiohub.org/). In addition to its core team, the HCA is poised to grow by community interaction, and states in its opening paper that "As with the Human Genome Project, a robust plan will best emerge from wide-ranging scientific discussions and careful planning".[2]
Thus, this project inserts itself among the wide-ranging scientific discussions to improve data - and knowledge - interoperability.

The highlight of "knowledge" in the last paragraph is meant to stress that raw data *per se* is not enough to turn the Atlas objectives into reality. There is a long way from raw datasets to commonly agreed scientific knowledge. And, ultimately, this long way is what allows humanity to take advantage of scientific endeavors. Currently, the gap between data and knowledge is mostly targeted via the writing and sharing of scientific manuscripts, the *de facto* currency of exchange of claims about the natural world. The Human Cell Atlas Publication Commitee reviews and selects publications that are directly part of the HCA. A set of publications is, thus, one of the major outputs of the whole endeavor.

The challenge that arises, thus, is one of managing a wealth of information and cast it into useful science. Experimental articles that analyze thousands of cells pose an overload of information alone. Ideally, we would like to understand, remember and make use of every statement produced by the HCA. As this goal is humanely impossible, we need to develop tools to make the knowledge interoperable with the aid of computers. At that point, the challenges of the HCA enter in resonance with the challenges of text-mining, biocuration and literature based discovery, which will be discussed in the chapter of this introduction.

Classification of cells into types

Given that a core goal of the Human Cell Atlas is to advance knowledge about *all* human cell types, [2] the definition of what a cell type is becomes important. Although a number of views exist [1,7,8,9,10,11,12,13,14,15,16,17,18], there is no formal, commonly agreed upon defintion of cell type. A 2017 article on the Human Cell Atlas mentions[10]:

"Descriptors such as 'cell type' and 'cell state' can be difficult to define at the moment. An integrative, systematic effort by many teams of scientists working together and bringing different expertise to the problem could dramatically sharpen our terminology, and revolutionize the way we see our cells, tissues and organs. We invite you to join the effort." The article highlights both the current gap in knowledge and the need of a community effort to work in that direction, in a direction that justifies the existence of the present work.

One consequence of a lack of a definition is that there is no commonly agreed number of cell types, and not even on an order of magnitude. As of November 2021, the leading answers in the Google Search Engine for the question "How many different cell types are found in the human body?" all point to around 200 different types (https://askabiologist.asu.edu/questions/human-cell-types, https://www.researchgate.net/post/How-many-cell-types-in-a-human-body-How-about-the-number-of-cell-cycles-in-each-species, https://www.kenhub.com/en/library/anatomy/types-of-cells-in-the-human-body), an estimate that is agreed upon by Bionumbers, a database of useful biological numbers [19] (https://bionumbers.hms.harvard.edu/bionumber.aspx?id=103626). A list of cell types in the adult human body on Wikipedia also amounts to around a couple hundred cell types [20, =List_of_distinct_cell_types_in_the_adult_human_body&oldid=1044853788]. However, the Cell Ontology has so far had catalogued 2,311 cell types of interest for the Human Cell Atlas as of June 2021 [21], increasing the estimate by at least one order of magnitude. Additionally, with an estimate of 37 trillion cells on average per human body [22] and an ever-increasing report of new cell types/clusters in single-cell transcriptomics ([23]), a precise estimate is not reasonable. In fact, the Human Cell Atlas project itself does not commit to any estimates of numbers of cell types, due to the sheer difficulty of estimating a number given current knowledge. (Aviv Regev; reply to question in the HCA conference)

Even though there is no agreement, different views on cell types are maturing. One core line of thought to define "cell type" is based on the cell type as an evolutionary unit defined by a Core Regulatory Complex (CoRC) of transcription factors. That definition enables the drawing of parallels, from the evolution of other biological entities (such as genes, proteins, and species) to cell types' evolution. Models of how multicellular life works greatly benefit from concepts such as "sister types" (cell types that diverged from a single ancestor), "cell type homology" (cell types in different species that share a common evolutionary origin), and "cell type convergence" (cell types that execute similar functions but which are not directly evolutionarily related) [24,25]

Another direction is based on the notion of attractors: regions of dynamical stability in a feature space, which might have different qualities. [26,27] In this theory, "basins of attraction" direct cell phenotypes, providing points in, say, a gene expression space towards which different cells "move" their expression programs. This dynamic view see each cell type corresponding to "a self-stabilizing regulatory program, which acts to maintain and restore the cell type-specific program of gene expression." [28] It alligns itself with dynamic systems theory, and some authors go as far as to say that "Lacking the idea of attractors we have no clear idea of what a cell type is." [29]

As much as different concepts of species coexist [30], our quest to define cell types may take various forms. The challenge of representing cell types in the context of evolution is conceptually different from representing cell types in biomedical experimentation. In that second direction, the groundwork of the Cell Ontology [31,32,33] and CELDA [34] and the contributions of the International Workshop on Cells in Experimental Life Sciences series [35,36] are notable.

Even though many sources of knowledge contribute to our knowledge about cell types [37], arguably single-cell transcriptomics is the workhorse for current efforts of the HUman cell Atlas, with an increasing amount of published studies using the methodology and of cells per study. [37] Current scRNA-seq data analyses often rely on unsupervised clustering of cells followed by assignment of cell-type labels to clusters. For the clustering, bioinformaticians tailor parameter sets to a target resolution, i.e., the level of detail used to detect cell identities. [38] [2] When the clustering is finished, the groups of cells are annotated with class labels, representing the underlying biology in a language we can understand. [39]

Instead of assigning expression gates from pre-defined markers, as is the standard for flow-cytometry analysis, single-cell RNA-seq analysis pipelines usually start from *de novo* clustering of cells followed by cluster annotation. [38] While it is clear that clusters and cell types are different concepts [38], often cluster labels are treated as cell types. There are a number of ways to cluster cells to find groups of similarity, but arguably the current default is dirived on the methodology proposed by PhenoGraph. [40] The protocol is to calculate the distances between cells in a reduced PCA space (with the number of dimensions chosen by the experimenters), followed by constructing a k-nearest-neighbours network, in which each cell is a node connected by *k* (another parameter) edges to other cells. Once the network is build, network modules (i.e. cell clusters) are commonly found using the Louvain algorithm, published in 2008 by researchers of the Université catholique de Louvain, in Belgium. [41] The cell clusters found by the PhenoGraph (or any other) algorithm are then labeled by domain experts, often based on genes differentially expressed on each cluster, so-called "markers". [38]

While it is possible to manually investigate the identities of which clusters, automatic methods have been developed to aid on the task. [39] One approach ("marker-based automatic annotation") bases itself on crossing clusters markers in the analyzed dataset with previous knowledge from databases like PanglaoDB [42] and CellMarker [43] [39]. Another approach (reference-based automatic cell annotation) relies on base, expert-annotated datasets as references from which labels are transferred to the dataset of interest. [39] Other methods bypass the clustering step and focus on labelling the individual cells, which avoids lumping dissimilar cells together, but require a high amount of reads per individual cells for it to be efficient. [39] A recent review and tutorial by Clarke et al [39] provides an extensive account of current techniques.

Of note, even though a range of methods is available, the vast majority of techniques and publications do not use standard identifiers for cell types. This is in contradiction with the acknowledgement by the community of the advantages of using identifiers the ad using standard identifiers, such as those provided by the Cell Ontology. [39] [44] [21] [45] [46] [47]. Nevertheless, projects that use Cell Ontology identifiers for single-cell RNA-seq data are appearing [48], including python and R packages (e.g. Besca [49], OnClass [50] and ontoProc[51]), data management projects and reference datasets, (e.g. Tabula Muris [52/] and Tabula Sapiens [53] Azimuth map [54/] and HubMap's ASCT+B Tables [55]) and annotation platforms (e.g. the Cell Annotation Platform [56] and CellTypist [57].

As elegantly put by Meehan et al [58] the Cell Ontology is a "manually constructed computer readable resource that links cell types by different relationships". it was first described in 2005 by Jonathan Bard, Seung Y Rhee† and Michael Ashburner [31] and was oriented at creating an "organism-independent classification of cells", following criteria that included function, histology, lineage and ploidy and providing "Cell-type unique identifiers (ID) that can be incorporated into any database holding cell-type-associated knowledge." It also had a didatic goal in itself, as the authors mention [31]: "It is designed to be useful in the sense that a researcher should be able to find, in a rapid and intuitive way, any cell type in any of the major model organisms and, having found it, learn a considerable amount about that cell type and its relationships to other biological objects." The collaborative project gradually evolved and changed its design and scope to fit new needs. By 2011, for example, a need for computable definitions for hematopoietic cell types lead to a sizeable advance in the number and quality of immune cell types represented in CL. [59] It also included the addidion of species-specific cell types to better handle marker-based definitions, which are usually given at the species level. [59] Further developments over the years included both technical improvements as well of the addition of new cell types, and by the time of the last official CL publication, in 2016, it contained approximately 2,200 classes. [45]
The Cell Ontology, currently, is growing as a resource for the Human Cell Atlas and in providing identifiers for cell types [48].

In conclusion, the advancement of our *formal* classification of cell types, such as in the Cell Ontology, represents a tangible goal of current cell-oriented large scale projects. While purely theoretical developments have their value, refining the cell type theory in the context of knowledge management arguably will have a influence direcly on how the products of the Human Cell Atlas will impact modern science. One reason is that formal systems enable automation of knowledge integration, and can feed intelligent systems that aid current research practices. In the following chapter, it will be discussed how computer-based knowledge processing can influence life-sciences research, as well as discuss techniques and platforms to advance the frontier.

Literature Based Discovery

The amount of scholar information vastly outnumbers what single researchers can fathom. Nevertheless, the gap between single individuals and the collectively body of knowledge has been widening in an accelerated fashion. The explosion in the number of published articles is leading to a "tsunami of knowlegde", flooding the scientific literature with rich information. That trend became spacially clear during the COVID-19 pandemic, when the huge amount of research published made keeping up with the literature pratically impossible. [61] At the same time, articles themselves are becoming denser, as high-throughput (and high-information) technologies like single-cell RNA-sequencing get cheaper and widely used. In practice, thus, too much of the knowledge generated remains unseen by any individual researcher, limiting the reach of science in general.

The technological advances, however, are no yet met by equivalent knowledge-handling systems. Mainstream scientific publication is, nowadays, barely readable by machines. Articles are written for human consumption, using ambiguous natural language and relying on implicit conventions. Tables and data rarely make use of technical standars, such as employing URIs (Uniform Resource Identifiers or encoding information in RDF formats. In fact, those standards and their acronyms are foreign for most life scientists (personal observations), despite being the *de jure* gold standard for data quality. [62/] Interconnecting biomedical knowledge is an open challenge of our century, and there is a large way to go before society can fully benefit from the sum of all knowledge we generate.

The scientific community has pursue solutions for this tsunami of information from many different angles. Narrative reviews, systematic reviews and textbooks compile and synthetize information, providing a layer of processing. Biocuration efforts go a step further and transform unstructured information into structured information in knowledgebases, such as UniProt [63] and PDB. Text-mining apply a range of Natural Language Processing tools to try and extract biological relations, or provide guidance for biocurators. Elaborate knowlegde networks, like the STRING database [64] and Wikidata[65], combine information from different sources.

The synthesis effort of literature mining is not only an exhibition of the scientific claims in the literature. Interconnected knowledge provides a way to discover new, implicit knowledge, by applying logical reasoning to a dataset. A field denominated Literature Based Discovery [66] dedicates itself to this challenge: make actual discoveries (or at least very strong hypothesis) using as material plainly the existing literature. [67] The textbook example of Literature Based Discovery is described by Don Swanson's ABC model: If A is related do B, and B is related to C, then A and C are indirectly related [68]. In a seminal paper, Swanson showed an hypothesis about using fish oil (A) to treat Raynauld's disease (C), demonstrating that even though the specialized fish-oil (A) literature had shown its association (AB) with a set of blood parameters (B), and the specialized Raynauld's disease literature had show its association (BC) with the same set of parameters (B), the AC link was never made in the literature, despite its seeming obviousness [68]

Modern advancements of literature-based discovery rely on Natural Language Processing, Machine Learning and Knowledge graphs to make inferences on literature knowledge. Word embeddings, for example, are leading inference of properties of compounds based on their shared neighbourhood of words (the words before and after their mentiongs) with known compounds, thus making use of latent knowledge in the body of knowledge. [69] Other, more explicit approaches, rely on extracted relations embedded in knowledge graphs. As an example, the discovery of new RNA-binding proteins related to Amyotrophic Lateral Sclerosis by analysis of the Watson Drug Discovery gene-disease network. [70]

Knowledge graphs have a set of characteristics that make then useful for Literature Based Discovery: they represent multiple relations, allow inferences on top of those relations, and provide human understandability at every step, allowing for a dialog between expert humans and computing systems. The field of biomedical ontologies explores that direction in depth, and the community is building many solutions, widely applicable for the biomedical sciences.

For the Human Cell Atlas Project (as presented in the chapter) to maximize its benefit for society, it is arguably important that its knowledge products are inserted into the main route of automated knowledge discovery . That implies a task of building knowledge graphs able to deal with it at all layers, including the generated data and metadata, its range of different protocols, and the purified knowledge projects that are enshrined in publications. Thus, the chapter will present challenges and paths for applying literature based discovery on a large scale and with sufficient flexibility to deal with the Human Cell Atlas.

Ontologies

The classification of biological concepts is at the core of biology. At least since the Aristotelian endeavours to group classes of animals, a good part of the scientific work is to capture concepts into knowledge systems [71]. Linnaeus' binomial system for naming species and Mendeleiev's periodic table are likely the two most famous classification systems, but are part of a much larger ecosystem of structuring scientific knowledge.

On the 20th century, the development of the analytical philosophy of Russel and Wittgenstein and their search for formalizations [72] gradually layed the foundations for the the logic of scientific descriptions. Karl Popper and his "The Logic of Scientific Discovery"[73] was heavily influenced by analytical philosophy, and the field is at the foundation of the "falseability" system of Popper. Less known among life scientists, Tarski's inquiries on what can be considered to be "true" [74] were also

The whole movement for formalization of knowledge progressed on the computational end, and at the late 20th century were at the root of the functioning of the World Wide Web, the advent of computational ontologies and large scale knowledge graphs. In this chapter, I will provide an overview of ontologies and knowledge graphs and their use in today's biomedical sciences, alongside its future prospects.

The OBO Foundry and biomedical ontologies

An ontology, as used here, is a formal computational representation of reality, which tries to represent each concept (and their relations) as precisely as possible. [71]
Constructing an ontology is a process of selecting and defining terms of interest, selecting and defining relationships of interest and making statements about reality using terms and relationships. The Gene Ontology is probably the most well known biomedical ontology; it describes (among other things) different classes of biological process, related to each_other by "is_a" and "part_of relations. [75] [76].

The Gene Ontology is part of a much larger effort to formalize concepts across biology: the Open Biomedical and Biological Ontologies (OBO) Foundry. [77] Created in 2007, the OBO Foundry is a hub of biomedical ontologies that sets guidelines for the design and construction of high-quality ontologies.

The initial OBO Foundry united several independent ontologies, like the Cell Ontology (CL), the Disease Ontology (DO) and the Protein Ontology (PRO) under a common framework towards interoperability. At the same time, the creation of the Relation Ontology (RO) provided a go-to point for relations in biology that could them be reused by different ontologies.

OWL and ontology languages

One of the OBO Principles for its ontologies is that they should be resolvable as a "syntactically valid OWL file using the RDF-XML syntax." (http://www.obofoundry.org/principles/fp-002-format.html). The OWL Web Ontology Language was introduced as a standard by the W3C consortium in 2004. OWL is not a programming language, as it does not instruct computers to perform actions, but an ontology language, which allows computerizable descriptions of the world. Furthermore, it is an umbrella ontology language that includes several languages with varying levels of expressivity. Generally, more expressive languages can represent more complex ideas, but make computations harder.

Regardless of the sublanguage used by ontology it must be resolvable to an RDF-XML file. RDF stands for Resource Description Framework, another W3C standard built around a graph-based data model (https://www.w3.org/TR/rdf11-concepts/). Statements in RDF are triples consisting of 2 nodes (a subject and an object) and an edge (a predicate) connecting the nodes. All nodes and edges are represented in RDFs by International Resource Identifiers (IRIs), and there are many ways to lay out those IRIs on a text file to represent triples. One of those layouts is the RDF-XML syntax, inspired by the XML markup language. Arguably, other syntaxes (interchangeable with RDF-XML) are easier to read for human. As an example of an RDF triple, here is how one would represent in the Turtle RDF Syntax, the notion that plasmacytoid dendritic cells are a type of dendritic cells:

```
http://purl.obolibrary.org/obo/CL_0000784  http://www.w3.org/2000/01/rdf-schema#subClassOf
http://purl.obolibrary.org/obo/CL_0000451 .
```

Where http://purl.obolibrary.org/obo/CL_0000784 and http://purl.obolibrary.org/obo/CL_0000451 are the unique IDs in the Cell Ontology for "plasmocytoid dendritic cells" and dentritic cells, respectively, and http://www.w3.org/2000/01/rdf-schema#subClassOf is the identifier for the "subclass of" relation as defined by the RDF schema.

A longer explanation of the details of OWL and RDF is outside the scope of this work. This brief introduction has a dual goal of introducing the architecture of formal representations and of demonstrating the complexity of the system. There is a high energy barrier to acquire the knowledge and the technical skills to engage in ontology building. That complexity might be one of the reasons why a very small fraction of the biomedical communities represents data with ontologies and an even smaller fraction engages with ontology building.

Wikidata

Even though the Semantic Web (which ontologies are a part of) spawned with promises of a revolution in the way knowledge is shared, it is still to be widely known outside the semantic engineering. Two recent projects are playing a particularly important role in bringing the Semantic Web to a wider audience has been receiving a boost of attention recently powered by two large projects: the Google Knowledge Graph and Wikidata.

The Google Knowledge Graph introduced the Semantic Web *de facto* in the daily life of users of Google. [78/]. Its underlying structure is similar to the triples in an ontology, but it is less concerned with being logically coherent, and does have strict semantics of a representation. In that way, Google Knowledge Graphs can feed on a variety of sources and not crash if there is some data modelling that, rigorously, could be inconsistent. Even though there is not a rigorous boundary between ontologies and knowledge graphs, one reasonable interpretation is that a knowledge graph may not be perfectly coherent, as long as it still can provide enough knowledge and reasoning for the approach of interest. While the lack of formal semantics limits reasoning and inference, the knowledge graphs are arguably easier to use, edit and and understand, and so provide an user friendly alternative for computable information with a lower entry barrier.

While the Google Knowledge Graph is widely used as a source of knowledge, it does not allow independent users to contribute with information. On the other hand, Wikidata, the collaborative knowledge graph of the Wikimedia foundation, allows users to contribute with classes and statements, in the same spirit of Wikipedia and share its "epistemic virtues, like power, speed and availability. [79] Its power is derived of its large community of contributors, closedly linked to the hugely successful Wikipedia. With a community of more than 20,000 active editors (https://www.wikidata.org/wiki/Wikidata:Statistics) and growing, it is able to cover a much wider number of concepts than any user individually. It is fast, because one does not need to install any software or ask for permissions to update it: any user can simply do it via a web interface. That speed makes it easier for newcomers to join and contribute, in contrast to OBO Foundry ontologies, which require extensive training on semantics and knowledge of Git/GitHub for contributions. Finally, the information on Wikidata is available via an user interface, via a SPARQL query service and as large, full-size database dumps, providing full extent reusability. The Wikidata model has been so sucessfull that Google decided to migrate its own knowledge base, Freebase, fully into Wikidata.[80]

The inner workings of Wikidata

Wikidata uses the same framework (RDF) that powers ontologies, and its model represents statements about the world in triples containing a subjects, a property and an object. [81] Its data model is serialized both in JSON and RDF. The data model contains 17 different datatypes, including, for example "Item", any entry on Wikidata that refers to "o a real-world object, concept, or event that is given an identifier in Wikidata" and "String", a "sequence of freely chosen characters interpreted as text". [82]. Knowledge is stored on Wikidata upon basic triples containing a subject (of type "Item"), a property and a value (which can be of any of the 17 types). As of November 2021, Wikidata contains more than 90 million data items [83] and more than 9000 properties that link them to values. As values often are other items, the database aquires a network format with labeled edges.

As can be seen in the example in 1, each the items in the database contain an item identifiers (Q followed by numbers). They also contain a label, a description and a list of aliases, which can be recorded in any of the more than 200 hundred languages, thereby making it a multilingual project. [84] Each item is decorated with statements, comprising of property-value pairs. These pairs can be further specified via qualifiers and references, which treats the full triple as the subject, adding metadata to it (a process called reification [85/#reification]). Qualifiers provide ways to extend the information on the triple, while references provide provenance, enabling users to judge the validity of the claims in the database.

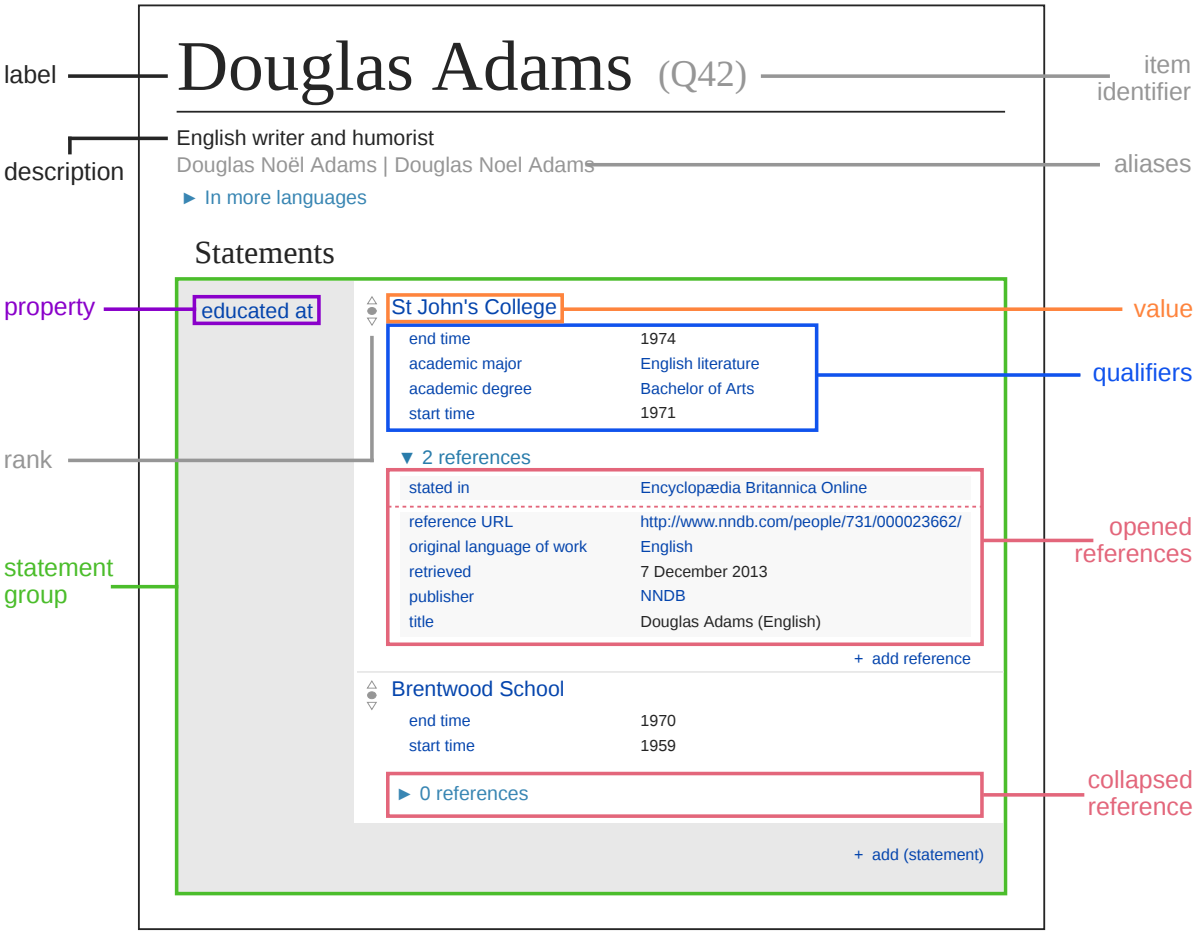


Figure 1: Wikidata's model for describing an item. Image released in Public Domain by Charlie Kritschmar.

All the information is available on a user interface, but its data is also available programatically in diverse formats, including as full JSON and RDF dumps, the MediaWiki API and a SPARQL endpoint. [86] A number of wrappers of such services are available in languages such as R [87] and python [88/]. A scheme of the data can be seen in 2, where each item is connected to a statement node via a property in the "p:" namespace, from which references and qualifiers are acessible. To facilitate basic usage, the namespace "wdt:" connects items to values directly, simplifying, for example, the writing of SPARQL queries.

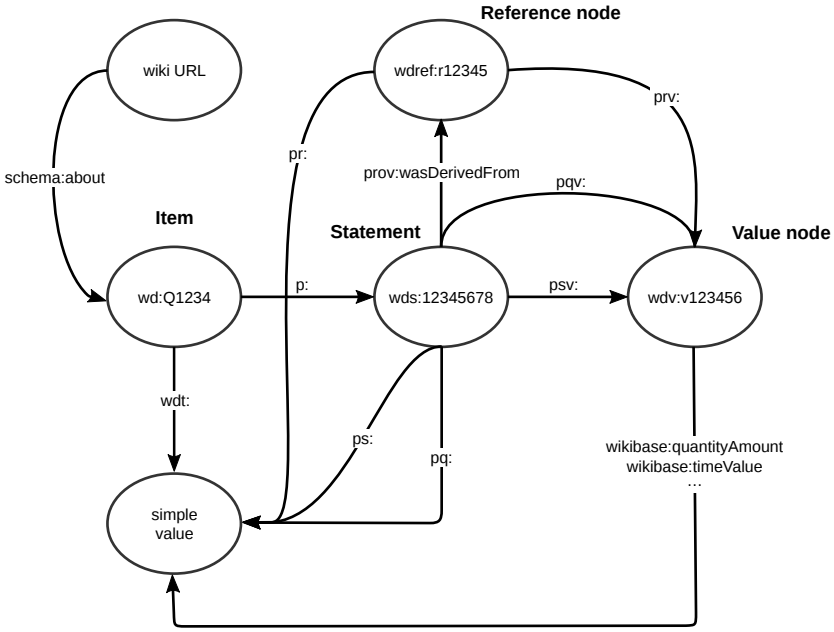


Figure 2: Wikidata's data model, scheme released under the CC-BY 4.0 license by Michael F. Schönitzer. It outlines the basic representation of statements, qualifiers and values in the Wikidata database

Information on Wikidata is released under a CC0 license, which enables full reuse of the data. [89] One of the major points of access and reuse of the information is the Wikidata Query Service [90/], a core resource of the community which enables live querying in the SPARQL language. [91] A number of services make use of embedded queries from the Wikidata Query Service [90/] to create interactive, live dashboards, for example Scholia [92/] aht the SARS-CoV-2 Query Book [93/]

Wikidata is not only acessible in different ways, but also writable in many ways. It provides a user-friendly, point-and-click interface for modifying the database, providing a low entry barriers for newcomers. It is also possible to semi-automatically reconcile spreadsheets to Wikidata items and use batch tools such as Open Refine [94] and Quickstatements [95], which enable batches on the magnitude of thousands of edits. For larger amounts of edits, it is possible to ask for bot permissions [96] and deploy systems that integrate big data sources. Bot edits are made via the WikiMedia API and are predominantly written via Python wrappers, such as Pywikibot [97] and the Wikidata Integrator. [98]

Wikidata as a knowledge graph for the life sciences

- Wikidata as a knowledge graph for the life sciences

Due to its privileged position inside the linked data ecosystem and its ease of write and query, Wikidata has been growing as a hub for interoperable data for the life sciences community. [65] [99] Even though Wikidata was created in 2013, the demand for a community-cured life sciences knowledge graph is clear at least since 2008 [100] [101] The Wikidata-like project proposed at the time was eventually discontinued, an example of the challenge of maintaining independent biomedical databases. [102] As Wikidata has a very large community, has stable funding and is at the core of modern technologies, like the Google Knowledge Graph [80] and Amazon's Alexa, [103/] it is virtually guaranteed that data in Wikidata will remain acessible for a long time, regardless of local funding schemes.

The Gene Wiki project [104] was likely the first large scale biomedical project to rely directly on the Wikipedia infrastructure for community curation. It provided a direction connection between the generalist community of Wikipedia and domain experts. The interplay of both communities is a topic of discussion and the opportunities and challenges were already discussed in NAR in 2012. [105]
Notably, Wikidata appeared chronologically after those efforts.
Notwithstandingly, the Gene Wiki research group has embraced the Wikidata environment for community biocuration and data interoperability [106] [107] [65] [108]. The information on Wikidata is still integrated to Wikipedias across multiple languages, often as source of information in Wikipedia's infoboxes.

Other projects outside the Gene Wiki innitiative also started using Wikidata as a platform for knowledge integration. A list of several projects that use Wikidata as part of their service to their community is given on table 1. There is movement exploring how Wikidata can be employed to the advance of Computational Biology, and how it can be integrated to current publication status quo. [109] In that direction, Wikidata is being developed as a platform for scholarly linked open data, particularly via the Scholia platform [110] [111],(https://scholia.toolforge.org/) which provides profiles of pre-templated SPARQL queries for entities like particular authors and articles (e.g. Scholia profile on Prof. Helder Nakaya: https://scholia.toolforge.org/author/Q42614737).

Table 1

During the COVID-19 pandemic, Wikidata has spawned as a hotspot for modelling information about the virus and the pandemic in real time. [112] [**wikidata:99196713?**] The general scope of the databasae allowed representation in a shared system of molecular, epidemiologic and socio-economic aspects of the pandemic. [112][113] Information curated in Wikidata was immediately available, feeding live dashboards and other applications based on SPARQL queries .[114] [115] [116] Additionally, as the information presented on Wikidata is multilingual and collaboratively edited, it presented itself as a resource for constructing structured vocabularies in non-english languages. [117]

In addition to its value as a structured database, Wikidata is tightly connected to Wikipedia. The gene identifiers in the context of Gene Wiki [106] are now fed to Wikipedias across languages, benefitting users directly. Additionally, gene expression information from the Bgee database [118] was added to Wikidata and connected to Wikipedia, which lead to a sizeable increase of the Bgee database. Currently, Wikipedia is one of the top 3 sources from which people access Bgee (personal communication with Tarcisio Farias, https://scholar.google.fr/citations?hl=fr&user=sB87J-cAAAAJ), thus leading to direct recognition for integrated bases. More generally, the connections of Wikidata and Wikipedia make it unique in the power of flowing knowledge back to human-acessed interfaces. In the words of Matthias Samwald [119] and colleagues "Wikidata could emerge as a community-backed and highly visible structured knowledge base of medical and biological information, bringing concepts and methodologies such as controlled taxonomies, Semantic Web / semantic technologies and ontologies into mainstream use."

In conclusion, Wikidata's unique position, robustness and guarantee of long term stability, prompts the need of works exploring new ways of integrating it to current knowledge management. Given the speed and breadth of the Human Cell Atlas, and the challenges of knowledge representation on cell types, this PhD work plans on discovering and adressing knowledge gaps on how Wikidata can play a role in organizing and disseminating the discoveries about all human cell types.

Objectives

- Study and refine theories of classes of cells within the constraints of ontologies and knowledge bases
- Provide a comprehensive list of currently described cell types on Wikidata
    - Develop a biocuration framework for the task of sharing information on Wikidata
    - Catalog as many cell types as possible, as a groundwork for future applications
- Devise ways to connect the Human Cell Atlas and other life-sciences products to Wikidata:
    - Craft wikidata relations ("properties") for making cell-type-related assertions
    - Write bots and scripts to reconcile data sources to Wikidata
- Provide proofs-of-concepts of how Wikidata integration can benefit the advancement of HCA

Methodology

This project's methodology resembles practical research-action practices [120]. Its goals of improving inteoperability of cell-type data implies a combination of action and research. Action in the form of active contributions to ontologies and knowledge-graphs, by getting involved and contributing to ongoing projects in the context of the Human Cell Atlas and knowledge management. Research in the 3 forms: - Philosophical investigation on the nature of knowledge representations of cell types, both in formal logic settings and in current academic practice - Applied investigations of database

integration and data quality in the context of Wikidata and biomedical ontologies - Data-driven biomedical research targeted at hypothesis generation and literature-based discovery using knowledge at the level of cell-type

All the research forms are intertwined with the improvement of knowledge management in biomedical sciences, with a focus on the Human Cell Atlas. The methods included the development and application of a framework for organized reading of the scientific literature, aimed at providing contact with the different facets of biocuration and Human Cell Atlas-related research.

Organized reading

In order to handle the literature reading necessary for this project, a framework was developed for reading. It consists of a set of Python scripts and a standard file structure. A file contains the list of articles to be read in markdown. Articles are represented as Wikidata QIDs which enables automatic information retrieval from Wikidata's structured systems. Articles were organized in 2 main sections, one for cell-type related articles and one for biocuration-related articles.

The Wikidata Bib system has a "pop" function, which creates a personalized note document in markdown for the first article of a section and (if possible) obtains the full text article from Unpaywall (https://unpaywall.org/). The note document contains a space for highlights, which can be copied from the original text and pasted in the markdown file. Notes and additional information are saved in a GitHub repository, and the structured information powers a live website with analytics on the users recent readings. The source code for Wikidata Bib is available at https://github.com/lubianat/wikidata_bib/tree/template and notes on my readings can currently be accessed at https://lubianat.github.io/wikidata_bib/.

Besides the technical aspects of Wikidata Bib, the organized reading methodology included a discipline step of continued reading, with a target of 1-2 papers per section per day. Based on Umberto Eco's suggestion on How to Write a Thesis [121] to develop a careful indexing system for literature, an index document was constructed containing the topics of interest for writing the thesis. The topics were added as plain text in the personalized note documents, allowing batch retrieval of articles of interest via the command line, using `grep` (https://en.wikipedia.org/w/index.php?title=Grep&oldid=1039541979).

Biocuration of cell classes for Wikidata

For each article about cell types read, cell types previously absent on Wikidata are added via a combination of curation in a Google Spreadsheet and a custom Python script (https://github.com/lubianat/wikidata_markers/tree/master/curation_of_classes).

Marker information was also recorded when explicitly mentioned and it will be added to Wikidata at a later step.

Annotation of Human Cell Atlas articles

Human Cell Atlas publications (https://www.humancellatlas.org/publications) were selected and abstracts were annotated as richfully as possible with Wikidata IDs using the hypothes.is annotation system (https://web.hypothes.is/). One article [2], describing the complete Human Cell Atlas project, was annotated in full. Annotations were retrieved via the hypothes.is API and processed with custom Python and R scripts (https://github.com/lubianat/ann/tree/main/hypothesis_parsing).

Wikidata updates

Wikidata is similar to a graph database, and is flexible enough to add new relations without need to change the underlying infrastructure.

Creation of new entities was done either manually in the Graphical User Interface (https://www.wikidata.org/wiki/Special:NewItem) or via custom python scripts combined with the Quickstatements tool (https://quickstatements.toolforge.org/#/) or the Wikidata Integrator python library (https://github.com/SuLab/WikidataIntegrator).

Properties, which link items to values, cannot be created at will and need to undergo community approval. Under the scope of this PhD project, we have gotten the community approval for a number of properties:

- entry receptor (https://www.wikidata.org/wiki/Property:P8339) used to link pathogens to their cellular entry receptors.
- Cell Ontology ID (https://www.wikidata.org/wiki/Property:P7963) used to link cell types to their IDs in the Cell Ontology
- has marker (https://www.wikidata.org/wiki/Property:P8872) used to link cell types to genes and proteins considered their markers
- derived from organism type (https://www.wikidata.org/wiki/Property:P9072) used to link cell lines to the taxon of the organism from which it was derived.

The property acceptance cycle takes at least one week and is completely open for opinions by any Wikidata user. All the information regarding the property proposal is available at https://www.wikidata.org/wiki/Wikidata:Property_proposal.

Cell Ontology participation

As part of the research-action process, I have joined the Cell Ontology working group

Data retrieval

- SPARQL queries

Data analysis

- Packages used in R and Python
- For interacting with Wikidata

Status of cell type info on Wikidata

Cell-disease network analysis

Preliminary Results

The concept of cell type

- Describe background
- Cell types, cell states and cell classes
- Levels of cell type information: archetype, senso stritu cell type, infratype and technotype.
- Infratypes and technotypes as theoretical innovations
- Current usage mixes archetypes and species-specific cell types
- Multi-level theory of conceptual modelling and a base level for biocuration of cell types

PanglaoDB integration to Wikidata

- The architecture of marker information on Wikidata
- Integration of information to the larger scope –> live updates by everyone
- Overview of the stats

Wikidata and the Cell Ontology interplay

Wikidata Bib and a professional system for biocuration

Introduction

- Accountants have Double-entry bookkeeping (https://en.wikipedia.org/wiki/Double-entry_bookkeeping), software developers have Test-driven development (https://en.wikipedia.org/wiki/Test-driven_development).

- Develop a professional way for coverage of large-scale revisions, inspired by Umberto Eco's How to Write a Thesis [121] adapted to the digital environment, using version-control and semantic links.

- Connect the reading framework with a Biocuration strategy to feed knowledge to Wikidata

Working of the system

Results

- Total number of cells on Wikidata
- Cells edited/added by me

Introduction

z

Concept of cell types

As an initial step of this PhD project, we decided to investigate the definition of "cell type" and how to shape a definition for knowledge management on Wikidata. The definition of the concept of "cell type" is currently a topic of debate by the biomedical community.[1,7,8,9,10,11,12,13,14,15,16,17]. Before we proceeded with the knowledge-graph formalizations via Wikidata, we dedicated time for a theoretical research on the concept of "cell type" in the context of knowledge representation. This line of research aligns itself with the groundwork of the Cell Ontology [31,32,33] and CELDA [34] and the contributions of the International Workshop on Cells in Experimental Life Sciences series [35,36].

In this period, we targeted the question: which cell type definition allows crafting coherent biological statements? The goal was to not say what cell types *are*, but what they can be for a consistent representation on an ontology or a knowledge graph, like Wikidata. We avoided the dissection of the differences between persistent classes of cells (often called "cell types") or the transient, fugacious classes of cells (often called "cell states") (see "Definition of cell identity" section in [122] for an example). Even though such a distinction is an essential topic for theoretical research, it is not required to represent formally biomedical experiments.

To facilitate communication among life scientists, in a preprint derived from this PHD project[123], we proposed, among other theoretical novelties, naming conventions for different cell types classes. Much of the literature mixes cell types in one species (e.g., when dealing with a cell type as an evolutionary unit) or multiple species (e.g., in the Cell Ontology). It is useful to distill these different concepts into names. Given the importance of the species' concept in biological classification [124], we derive a species-centric view on the naming of classes of cell types. The four classes (Figure 3) we propose are as follows:

- archetypes, for when the taxonomic scope of the type is beyond the level of species; for example, "mammal neutrophils."
- *sensu stricto* cell types, for when the taxonomic scope of the type corresponds to a single species; for example, *Mus musculus* neutrophils."
- infratypes, for when the taxonomic scope is below the level of species; for example, considering the mouse strain "C57BL/6J", "neutrophils from C57BL/6J mice".
- technotypes, for specific, experimentally defined cell types that harbor in their definition the precise conditions of the cells sampled; "2-month-old male C57BL/6J, Ly-6G^{+} CD11b^{+} M-CSF R^{-} CD244^{-} neutrophils".

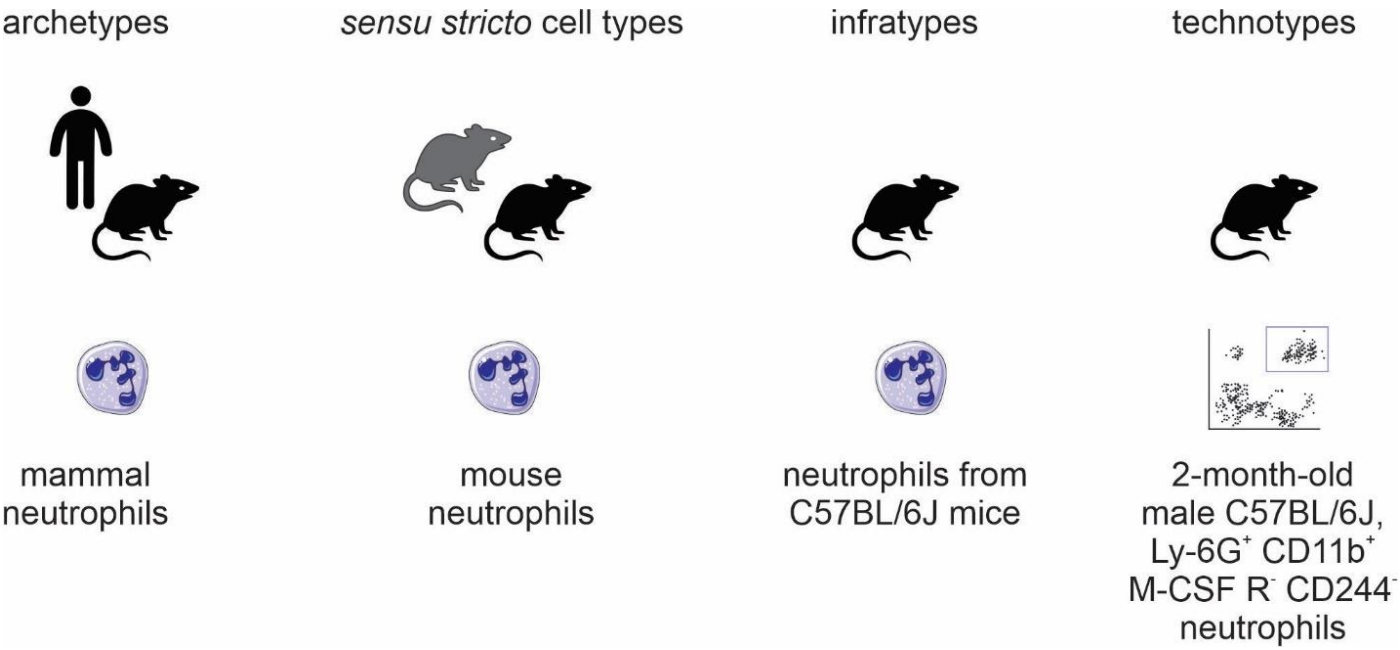


Figure 3: Names for classes of cell types.

The 4 different categories of cell types help us to better organize the knowledge about cell types. Even though individual articles and databases often have species-neutral names, the information often comes from experiments with a single strain of a single species. Two articles might call by the same name cells that come from different animals, or were selected by different protocols. Large scale knowledge management requires an organized way of representing those details.

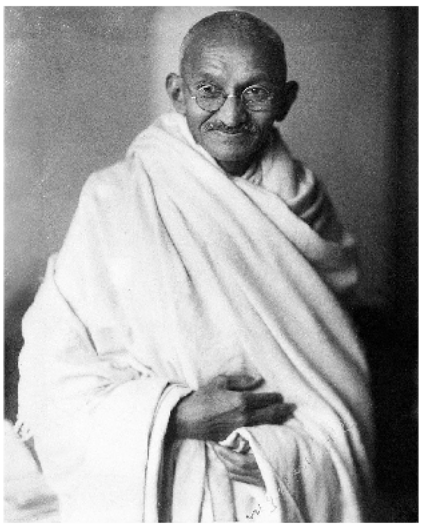
The division between archetypes and *sensu stricto* cell types is of special importance in the context of biocuration and annotation of data. Associations like the HUGO Gene Nomenclature Committee and UniProt organize names and identifiers for genes and proteins in single species. Thus, if we want to annotate marker genes, for example, we need to associate them to a species-specific cell type (a *sensu stricto* cell type) instead of the more vague association to a species-neutral type. That might seem obvious, but current standards still use identifiers that are species-neutral (e.g. in the reference HuBMAP app; https://azimuth.hubmapconsortium.org/references/)

The ontological discussion on the classes of cell types, thus, extends the current state-of-the-art and introduce new ways to organize our knowledge about cells. Notably, the technotype and the infratype are, currently, mostly theoretical constructs and almost no resources deal with cell types at the level of strains or below. The division of archetypes and *sensu stricto* cell types, on the other hand, was already instrumental for the integration of the Panglao database of cell markers to Wikidata, described in the next session.

- Add bits about multilevel theory and pragmatic defintion

realm

kingdom of individuals



Mahatma Gandhi

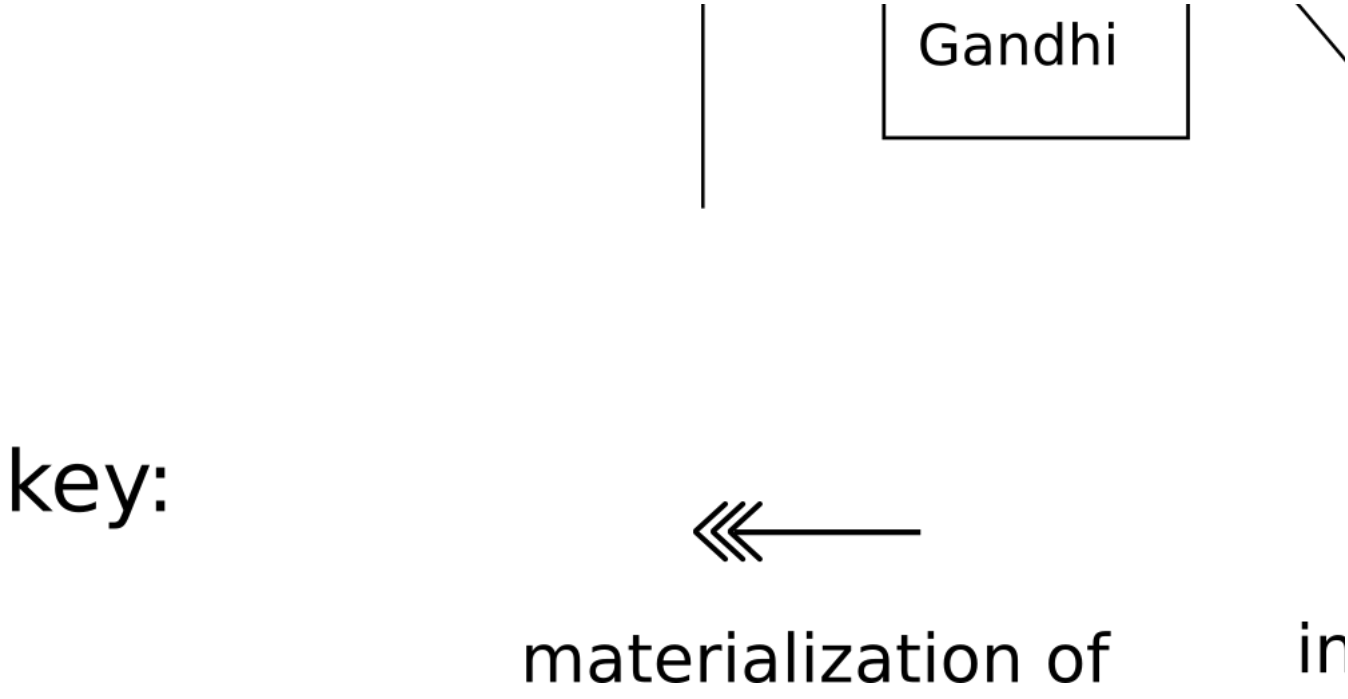
B.

material realm

th

kingdom of individuals

zygote that became Mahatma



Preliminary Results

Fcoex updates

The concept of cell type

- Describe background
- Cell types, cell states and cell classes
- Levels of cell type information: archetype, senso stritu cell type, infratype and technotype.
- Infratypes and technotypes as theoretical innovations
- Current usage mixes archetypes and species-specific cell types
- Annotation of HCA articles for grasping the use of different concepts in the context of HCA

Next steps

- Improve formalization of cell types in connection with the biomedical semantics community

HCA

- “Sky dive” approach: hand annotation of all abstracts and the core Human Cell Atlas paper
- Benefits of using a single ontology that anyone can edit (new terms and speed of science)
- Figure: The different concepts in use by the HCA paper
- Figure: The different concepts in use by the different HCA papers
- Discussion
- Information by HCA and related efforts is already targeted by biocurators. PanglaoDB is one of these resources etc etc

Next steps

- Mature the annotation system into a curation tool (based on ANN, perhaps reuse figure)
- Explore the use of SciSpacy and natural language processing for making it easier

Cell Ontology

PanglaoDB integration to Wikidata

- The architecture of marker information on Wikidata
- Integration of information to the larger scope –> live updates by everyone
- Overview of the stats

Preliminary Results

Fcoex updates

The concept of cell type

- Describe background
- Cell types, cell states and cell classes

- Levels of cell type information: archetype, senso stritu cell type, infratype and technotype.
- Infratypes and technotypes as theoretical innovations
- Current usage mixes archetypes and species-specific cell types
- Annotation of HCA articles for grasping the use of different concepts in the context of HCA

Next steps

- Improve formalization of cell types in connection with the biomedical semantics community

HCA

- “Sky dive” approach: hand annotation of all abstracts and the core Human Cell Atlas paper
- Benefits of using a single ontology that anyone can edit (new terms and speed of science)
- Figure: The different concepts in use by the HCA paper
- Figure: The different concepts in use by the different HCA papers
- Discussion
- Information by HCA and related efforts is already targeted by biocurators. PanglaoDB is one of these resources etc etc

Next steps

- Mature the annotation system into a curation tool (based on ANN, perhaps reuse figure)
- Explore the use of SciSpacy and natural language processing for making it easier

Cell Ontology

PanglaoDB integration to Wikidata

- The architecture of marker information on Wikidata
- Integration of information to the larger scope –> live updates by everyone
- Overview of the stats

Additional Works

fcoex

Cell Ontology - Minimal Information About a New Cell Type

Cellosaurus and Wikidata

Complex Portal and Wikidata

WikiProject ELIXIR

Systematic Reviews and publishing of intermediary tables

Academic Curriculum

University course

Awards and Participation in events (?)

- ISCB 2021
- BioHack EU 2021

References

1. **An era of single-cell genomics consortia**
Yoshinari Ando, Andrew T Kwon, Jay W Shin
Experimental and Molecular Medicine (2020-09-15) https://www.wikidata.org/wiki/Q99418649
DOI: 10.1038/s12276-020-0409-x

2. **The Human Cell Atlas.**
Aviv Regev, Sarah Teichmann, Eric Lander, Amir Giladi, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, Peter Campbell, Piero Carninci, Menna R Clatworthy, ... Human Cell Atlas Meeting Participants
eLife (2017-12-05) https://www.wikidata.org/wiki/Q46368626
DOI: 10.7554/elife.27041

3. **The Human Cell Atlas and equity: lessons learned**
Partha P Majumder, Musa M Mhlanga, Alex K Shalek
Nature Medicine (2020-10-01) https://www.wikidata.org/wiki/Q100491106
DOI: 10.1038/s41591-020-1100-4

4. **The Human Cell Atlas White Paper**
Aviv Regev, Sarah Teichmann, Orit Rozenblatt-Rosen, Michael JT Stubbington, Kristin Ardlie, Amir Giladi, Paola Arlotta, Gary D Bader, Christophe Benoist, Moshe Biton, ... Human Cell Atlas Organizing Committee
(2018-10-11) https://www.wikidata.org/wiki/Q104450645

5. **Everyone needs a data-management plan**
Nature
(2018-03-15) https://www.wikidata.org/wiki/Q56524391
DOI: 10.1038/d41586-018-03065-z

6. **About the Data Coordination Platform**
HCA Data Portal
https://data.humancellatlas.org/about/

7. **What Is Your Conceptual Definition of "Cell Type" in the Context of a Mature Organism?**
Paul Blainey, Hans Clevers, Cole Trapnell, Ed Lein, Emma Lundberg, Alfonso Martinez Arias, Joshua R Sanes, Jay Shendure, James Eberwine, Junhyong Kim, ... Mathias Uhlén
Cell systems (2017-03-01) https://www.wikidata.org/wiki/Q87649649
DOI: 10.1016/j.cels.2017.03.006

8. **A periodic table of cell types**
Bo Xia, Itai Yanai
Development (2019-06-15) https://doi.org/ggctwf
DOI: 10.1242/dev.169854 · PMID: 31249003 · PMCID: PMC6602355

9. **Exciting times to study the identity and evolution of cell types**
Maria Sachkova, Pawel Burkhardt
Development (2019-09-15) https://doi.org/ghdb9v
DOI: 10.1242/dev.178996 · PMID: 31537583

10. **The Human Cell Atlas: from vision to reality.**
Orit Rozenblatt-Rosen, Michael JT Stubbington, Aviv Regev, Sarah Teichmann
Nature (2017-10-01) https://www.wikidata.org/wiki/Q47565008
DOI: 10.1038/550451a

11. **Human Cell Atlas and cell-type authentication for regenerative medicine**
Yulia Panina, Peter Karagiannis, Andreas Kurtz, Glyn N Stacey, Wataru Fujibuchi
Experimental and Molecular Medicine (2020-09-15) https://www.wikidata.org/wiki/Q99418657
DOI: 10.1038/s12276-020-0421-1

12. **A community-based transcriptomics classification and nomenclature of neocortical cell types**
Rafael Yuste, Michael J Hawrylycz, Nadia Aalling, Argel Aguilar-Valles, Detlev Arendt, Rubén Armañanzas, Giorgio A Ascoli, Concha Bielza, Vahid Bokharaie, Tobias B Bergmann, ... Ed S Lein
Nature Neuroscience (2020-08-24) https://www.wikidata.org/wiki/Q98665291
DOI: 10.1038/s41593-020-0685-8

13. **The evolving concept of cell identity in the single cell era**
Samantha A Morris
Development (2019-06-27) https://www.wikidata.org/wiki/Q93086971
DOI: 10.1242/dev.169748

14. **Implications of Epigenetic Variability within a Cell Population for "Cell Type" Classification**
Inna Tabansky, Joel Stern, Donald W Pfaff
Frontiers in Behavioral Neuroscience (2015-12-16) https://www.wikidata.org/wiki/Q26770736
DOI: 10.3389/fnbeh.2015.00342

15. **Geometry of the Gene Expression Space of Individual Cells**
Yael Korem, Pablo Szekely, Yuval Hart, Hila Sheftel, Jean Hausser, Avi Mayo, Michael E Rothenberg, Tomer Kalisky, Uri Alon
PLOS Computational Biology (2015-07-10) https://www.wikidata.org/wiki/Q35688096
DOI: 10.1371/journal.pcbi.1004224

16. **Evolution of Cellular Differentiation: From Hypotheses to Models**
Pedro Márquez-Zacarías, Rozenn M Pineau, Marcella Gomez, Alan Veliz-Cuba, David Murrugarra, William C Ratcliff, Karl J Niklas
Trends in Ecology & Evolution (2020-08-20) https://www.wikidata.org/wiki/Q98633613

DOI: 10.1016/j.tree.2020.07.013

17. **Inferring cell type innovations by phylogenetic methods-concepts, methods, and limitations**
Koryu Kin, Koryu Kin
Journal of Experimental Zoology. Part B: Molecular and Developmental Evolution (2015-10-14) https://www.wikidata.org/wiki/Q40436539
DOI: 10.1002/jez.b.22657

18. **Towards a pragmatic definition of cell type**
Tiago Lubiana, Helder Nakaya
(2021-01-04) https://www.wikidata.org/wiki/Q108723646
DOI: 10.22541/au.160979530.02627436/v1

19. **BioNumbers--the database of key numbers in molecular and cell biology**
Ron Milo, Paul Jorgensen, Uri Moran, Griffin M Weber, Michael Springer
Nucleic Acids Research (2010-01-01) https://www.wikidata.org/wiki/Q24643881
DOI: 10.1093/nar/gkp889

20. **Wikipedia, the free encyclopedia** https://en.wikipedia.org/wiki/Main_Page

21. **Cell types and ontologies of the Human Cell Atlas**
David Osumi-Sutherland, Chuan Xu, Maria C Keays, Peter V Kharchenko, Aviv Regev, Ed S Lein, Sarah Teichmann
(2021-06-28) https://www.wikidata.org/wiki/Q107373831

22. **An estimation of the number of cells in the human body.**
Eva Bianconi, Allison Piovesan, Federica Facchin, Alina Beraudi, Raffaella Casadei, Flavia Frabetti, Lorenza Vitale, Maria Chiara Pelleri, Simone Tassani, Francesco Piva, … Silvia Canaider
Annals of Human Biology (2013-07-05) https://www.wikidata.org/wiki/Q34037445
DOI: 10.3109/03014460.2013.807878

23. **A curated database reveals trends in single-cell transcriptomics**
Valentine Svensson, Eduardo da Veiga Beltrame, Lior Pachter
Database (2020-11-01) https://www.wikidata.org/wiki/Q103034964
DOI: 10.1093/database/baaa073

24. **The evolution of cell types in animals: emerging principles from molecular studies.**
Detlev Arendt
Nature reviews. Genetics (2008-11) https://www.ncbi.nlm.nih.gov/pubmed/18927580
DOI: 10.1038/nrg2416 · PMID: 18927580

25. **The origin and evolution of cell types**
Detlev Arendt, Jacob M Musser, Clare VH Baker, Aviv Bergman, Connie Cepko, Douglas H Erwin, Mihaela Pavlicev, Gerhard Schlosser, Stefanie Widder, Manfred D Laubichler, Günter P Wagner
Nature Reviews Genetics (2016-11-07) https://doi.org/f9b62x
DOI: 10.1038/nrg.2016.127 · PMID: 27818507

26. **Stem cell states, fates, and the rules of attraction.**
Tariq Enver, Martin Pera, Carsten Peterson, Peter W Andrews
Cell Stem Cell (2009-05-01) https://www.wikidata.org/wiki/Q37475461
DOI: 10.1016/j.stem.2009.04.011

27. **Theory of cell fate**
Michael J Casey, Patrick S Stumpf, Ben D MacArthur
Wiley interdisciplinary reviews. Systems biology and medicine (2019-12-12) https://www.wikidata.org/wiki/Q91908361
DOI: 10.1002/wsbm.1471

28. **Perspectives on defining cell types in the brain**
Eran A Mukamel, John Ngai
Current Opinion in Neurobiology (2018-12-06) https://www.wikidata.org/wiki/Q90361677
DOI: 10.1016/j.conb.2018.11.007

29. **Ensembles, dynamics, and cell types: Revisiting the statistical mechanics perspective on cellular regulation**
Stefan Bornholdt, Stuart Kauffman
Journal of Theoretical Biology (2019-01-31) https://www.wikidata.org/wiki/Q91316993
DOI: 10.1016/j.jtbi.2019.01.036

30. **Species Concepts and Species Delimitation**
Kevin De Queiroz
Systematic Biology (2007-12) https://doi.org/c34kzf
DOI: 10.1080/10635150701701083 · PMID: 18027281

31. **An ontology for cell types**
Jonathan Bard, Sue Rhee, Michael Ashburner
Genome Biology (2005-01-01) https://www.wikidata.org/wiki/Q21184168
DOI: 10.1186/gb-2005-6-2-r21

32. **Logical Development of the Cell Ontology**
Terrence F Meehan, Anna Maria Masci, Amina Abdulla, Lindsay G Cowell, Judith A Blake, Christopher J Mungall, Alexander D Diehl
BMC Bioinformatics (2011-01-05) https://doi.org/c7kw6x
DOI: 10.1186/1471-2105-12-6 · PMID: 21208450 · PMCID: PMC3024222

33. **The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability**
Alexander D Diehl, Terrence F Meehan, Yvonne M Bradford, Matthew H Brush, Wasila M Dahdul, David S Dougall, Yongqun He, David Osumi-Sutherland, Alan Ruttenberg, Sirarat Sarntivijai, … Christopher J Mungall
Journal of Biomedical Semantics (2016-07-04) https://doi.org/gg99b9

DOI: 10.1186/s13326-016-0088-7 · PMID: 27377652 · PMCID: PMC4932724

34. **CELDA -- an ontology for the comprehensive representation of cells in complex systems**
Stefanie Seltmann, Harald Stachelscheid, Alexander Damaschun, Ludger Jansen, Fritz Lekschas, Jean-Fred Fontaine, Throng Nghia Nguyen-Dobinsky, Ulf Leser, Andreas Kurtz
BMC Bioinformatics (2013-07-17) https://www.wikidata.org/wiki/Q21284308
DOI: 10.1186/1471-2105-14-228

35. **Cells in experimental life sciences - challenges and solution to the rapid evolution of knowledge**
Sirarat Sarntivijai, Alexander D Diehl, Yongqun He
BMC Bioinformatics (2017-12-21) https://doi.org/gg99b7
DOI: 10.1186/s12859-017-1976-2 · PMID: 29322916 · PMCID: PMC5763506

36. **Cells in ExperimentaL Life Sciences (CELLS-2018): capturing the knowledge of normal and diseased cells with ontologies**
Sirarat Sarntivijai, Yongqun He, Alexander D Diehl
BMC Bioinformatics (2019-04-25) https://doi.org/gg99b8
DOI: 10.1186/s12859-019-2721-9 · PMID: 31272374 · PMCID: PMC6509796

37. **Scaled, high fidelity electrophysiological, morphological, and transcriptomic cell characterization**
Brian R Lee, Agata Budzillo, Kristen Hadley, Jeremy A Miller, Tim Jarsky, Katherine Baker, DiJon Hill, Lisa Kim, Rusty Mann, Lindsay Ng, … Jim Berg
eLife (2021-08-13) https://www.wikidata.org/wiki/Q109717199
DOI: 10.7554/elife.65482

38. **Current best practices in single-cell RNA-seq analysis: a tutorial**
Malte D Luecken, Fabian J Theis
Molecular Systems Biology (2019-06-19) https://www.wikidata.org/wiki/Q64974172
DOI: 10.15252/msb.20188746

39. **Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods**
Zoe A Clarke, Tallulah Andrews, Jawairia Atif, Delaram Pouyabahar, Brendan T Innes, Sonya A MacParland, Gary D Bader
Nature Protocols (2021-05-24) https://www.wikidata.org/wiki/Q107158224
DOI: 10.1038/s41596-021-00534-0

40. **Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis**
Jacob H Levine, Erin F Simonds, Sean C Bendall, Kara L Davis, El-ad D Amir, Michelle D Tadmor, Oren Litvin, Harris G Fienberg, Astraea Jager, Eli R Zunder, … Garry P Nolan
Cell (2015-06-18) https://www.wikidata.org/wiki/Q30975629
DOI: 10.1016/j.cell.2015.05.047

41. **Fast unfolding of communities in large networks**
Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre
Journal of Statistical Mechanics: Theory and Experiment (2008-10-09) https://www.wikidata.org/wiki/Q29305711
DOI: 10.1088/1742-5468/2008/10/p10008

42. **PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data**
Oscar Franzén, Li-Ming Gan, Johan LM Bjorkegren
Database (2019-01-01) https://www.wikidata.org/wiki/Q63664483
DOI: 10.1093/database/baz046

43. **CellMarker: a manually curated resource of cell markers in human and mouse**
Xinxin Zhang, Yujia Lan, Jinyuan Xu, Fei Quan, Erjie Zhao, Chunyu Deng, Tao Luo, Liwen Xu, Gaoming Liao, Min Yan, … Yun Xiao
Nucleic Acids Research (2019-01-01) https://www.wikidata.org/wiki/Q56984510
DOI: 10.1093/nar/gky900

44. **Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data**
Julie A McMurry, Nick Juty, Niklas Blomberg, Tony Burdett, Tom Conlin, Nathalie Conte, Melanie Courtot, John Deck, Michel Dumontier, Donal K Fellows, … Helen Parkinson
PLOS Biology (2017-06-29) https://www.wikidata.org/wiki/Q33037209
DOI: 10.1371/journal.pbio.2001414

45. **The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability.**
Alexander D Diehl, Terrence F Meehan, Yvonne M Bradford, Matthew H Brush, Wasila M Dahdul, David S Dougall, Yongqun He, David Osumi-Sutherland, Alan Ruttenberg, Sirarat Sarntivijai, … Christopher J Mungall
Journal of Biomedical Semantics (2016-07-04) https://www.wikidata.org/wiki/Q36067763
DOI: 10.1186/s13326-016-0088-7

46. **Cell type discovery using single-cell transcriptomics: implications for ontological representation**
Brian D Aevermann, Mark Novotny, Trygve E Bakken, Jeremy A Miller, Alexander D Diehl, David Osumi-Sutherland, Roger S Lasken, Ed S Lein, Richard H Scheuermann
Human Molecular Genetics (2018-05-01) https://www.wikidata.org/wiki/Q52625486
DOI: 10.1093/hmg/ddy100

47. **Cell ontology in an age of data-driven cell classification.**
David Osumi-Sutherland, David Osumi-Sutherland
BMC Bioinformatics (2017-12-21) https://www.wikidata.org/wiki/Q49192555
DOI: 10.1186/s12859-017-1980-6

48. **Cell type ontologies of the Human Cell Atlas**
David Osumi-Sutherland, Chuan Xu, Maria Keays, Adam P Levine, Peter V Kharchenko, Aviv Regev, Ed Lein, Sarah Teichmann
Nature Cell Biology (2021-11-01) https://www.wikidata.org/wiki/Q109755180
DOI: 10.1038/s41556-021-00787-7

49. **Besca, a single-cell transcriptomics analysis toolkit to accelerate translational research**

Sophia Clara Mädler, Alice Julien-Laferriere, Luis Wyss, Miroslav Phan, Albert SW Kang, Eric Ulrich, Roland Schmucki, Jitao David Zhang, Martin Ebeling, Laura Badi, … Klas Hatje
bioRxiv (2020-08-12) https://www.wikidata.org/wiki/Q104450593
DOI: 10.1101/2020.08.11.245795

50. **Leveraging the Cell Ontology to classify unseen cell types**
Sheng Wang, Angela Oliveira Pisco, Aaron McGeever, Maria Brbić, Marinka Žitnik, Spyros Darmanis, Jure Leskovec, Jim Karkanias, Russ Altman
Nature Communications (2021-09-21) https://www.wikidata.org/wiki/Q108929315
DOI: 10.1038/s41467-021-25725-x

51. **ontoProc: processing of ontologies of anatomy, cell lines, and so on** https://www.wikidata.org/wiki/Q101074371

52. **Tabula Muris** https://tabula-muris.ds.czbiohub.org/

53. **Tabula Sapiens** https://tabula-sapiens-portal.ds.czbiohub.org/celltypes

54. **Azimuth** https://azimuth.hubmapconsortium.org/

55. **Construction and Usage of a Human Body Common Coordinate Framework Comprising Clinical, Semantic, and Spatial Ontologies**
Katy Börner, Ellen Quardokus, Bruce WHerr II, Leonard E Cross, Elizabeth G Record, Yingnan Ju, Andreas D Bueckle, James P Sluka, Jonathan C Silverstein, Kristen M Browne, … Griffin M Weber
(2020-07-28) https://www.wikidata.org/wiki/Q109755184

56. **Cell Annotation Platform | Coming Soon** http://celltype.info/

57. **Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture across the human body**
Conde C Domínguez, Tomás Gomes, Lorna B Jarvis, C Xu, SK Howlett, DB Rainbow, Ondrej Suchanek, Hamish W King, Lira Mamanova, Krzysztof Polański, … Sarah Teichmann
(2021-04-28) https://www.wikidata.org/wiki/Q107363182
DOI: 10.1101/2021.04.28.441762

58. **Ontology based molecular signatures for immune cell types via gene expression analysis**
Terrence F Meehan, Nicole Vasilevsky, Christopher J Mungall, David S Dougall, Melissa Haendel, Judith A Blake, Alexander D Diehl
BMC Bioinformatics (2013-08-30) https://www.wikidata.org/wiki/Q34978215
DOI: 10.1186/1471-2105-14-263

59. **Logical development of the cell ontology**
Terrence F Meehan, Anna Maria Masci, Amina Abdulla, Lindsay G Cowell, Judith A Blake, Christopher J Mungall, Alexander D Diehl
BMC Bioinformatics (2011-01-05) https://www.wikidata.org/wiki/Q33786317
DOI: 10.1186/1471-2105-12-6

60. **Scientists are drowning in COVID-19 papers. Can new tools keep them afloat?** https://www.science.org/content/article/scientists-are-drowning-covid-19-papers-can-new-tools-keep-them-afloat

61. **How a torrent of COVID science changed research publishing — in seven charts**
Holly Else
Nature (2020-12-16) https://www.nature.com/articles/d41586-020-03564-y

62. **5-star Open Data** http://5stardata.info/en/

63. **UniProt: the universal protein knowledgebase in 2021**
UniProt Consortium
Nucleic Acids Research (2020-11-25) https://www.wikidata.org/wiki/Q102383737
DOI: 10.1093/nar/gkaa1100

64. **The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets**
Damian Szklarczyk, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, Sampo Pyysalo, Nadezhda Tsankova Doncheva, Marc Legeay, Tao Fang, Peer Bork, … Christian von Mering
Nucleic Acids Research (2020-11-25) https://www.wikidata.org/wiki/Q102383784
DOI: 10.1093/nar/gkaa1074

65. **Wikidata as a knowledge graph for the life sciences**
Andra Waagmeester, Gregory Stupp, Sebastian Burgstaller-Muehlbacher, Benjamin M Good, Malachi Griffith, Obi Griffith, Kristina Hanspers, Henning Hermjakob, Toby Hudson, Kevin Hybiske, … Andrew I Su
eLife (2020-03-17) https://www.wikidata.org/wiki/Q87830400
DOI: 10.7554/elife.52614

66. **Literature-based discovery - Wikipedia** https://en.wikipedia.org/wiki/Literature-based_discovery

67. **Literature Based Discovery: models, methods, and trends.**
MSSam Henry, Bridget T McInnes
Journal of Biomedical Informatics (2017-08-21) https://www.wikidata.org/wiki/Q38371706
DOI: 10.1016/j.jbi.2017.08.011

68. **Online tools to support literature-based discovery in the life sciences.**
Marc Weeber, Marc Weeber, Jan A Kors, Jan A Kors, Barend Mons
Briefings in Bioinformatics (2005-09-01) https://www.wikidata.org/wiki/Q36280460
DOI: 10.1093/bib/6.3.277

69. **Unsupervised word embeddings capture latent knowledge from materials science literature**
Vahe Tshitoyan, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A Persson, Gerbrand Ceder, Anubhav Jain
Nature (2019-07-03) https://www.wikidata.org/wiki/Q91595456
DOI: 10.1038/s41586-019-1335-8

70. **Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis.**
Nadine Bakkar, Tina Kovalik, Ileana Lorenzini, Scott Spangler, Alix Lacoste, Kyle Sponaugle, Philip Ferrante, Elenee Argentinis, Rita Sattler, Robert Bowser
Acta Neuropathologica (2017-11-13) https://www.wikidata.org/wiki/Q47406275
DOI: 10.1007/s00401-017-1785-8

71. **Ontologies for the life sciences**
Steffen Schulze-Kremer, Barry Smith
(2005-11-15) https://www.wikidata.org/wiki/Q105870680
DOI: 10.1002/047001153x.g408213

72. **The Philosophy of Logical Atomism, Lecture 1: Facts and Propositions** https://www.wikidata.org/wiki/Q105105637

73. **Logik der Forschung**
Karl Popper
(1934-01-01) https://www.wikidata.org/wiki/Q1868040

74. **The semantic conception of truth: and the foundations of semantics**
Alfred Tarski
Philosophy and Phenomenological Research (1944-03-01) https://www.wikidata.org/wiki/Q106090790
DOI: 10.2307/2102968

75. **The Gene Ontology resource: enriching a GOld mine**
Gene Ontology Consortium
Nucleic Acids Research (2020-12-08) https://www.wikidata.org/wiki/Q104130127
DOI: 10.1093/nar/gkaa1113

76. **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**
M Ashburner, CA Ball, Judith A Blake, David Botstein, H Butler, JMichael Cherry, AP Davis, K Dolinski, Selina S Dwight, JT Eppig, … Gavin Sherlock
Nature Genetics (2000-05-01) https://www.wikidata.org/wiki/Q23781406
DOI: 10.1038/75556

77. **The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration**
Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J Mungall, … Suzanna Lewis
Nature Biotechnology (2007-11-01) https://www.wikidata.org/wiki/Q19671692
DOI: 10.1038/nbt1346

78. **Introducing the Knowledge Graph: things, not strings**
Google
(2012-05-16) https://blog.google/products/search/introducing-knowledge-graph-things-not/

79. **Toward an epistemology of Wikipedia**
Don Fallis
Journal of the Association for Information Science and Technology (2008-08-01) https://www.wikidata.org/wiki/Q101955295
DOI: 10.1002/asi.20870

80. **From Freebase to Wikidata: The Great Migration**
Thomas Pellissier Tanon, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, Lydia Pintscher
Proceedings of the 25th International Conference on World Wide Web (2016-01-01) https://www.wikidata.org/wiki/Q24074986
DOI: 10.1145/2872427.2874809

81. **Wikibase/DataModel - MediaWiki** https://www.mediawiki.org/wiki/Wikibase/DataModel

82. **Help:Data type - Wikidata** https://www.wikidata.org/wiki/Help:Data_type

83. **Wikidata:Statistics - Wikidata** https://www.wikidata.org/wiki/Wikidata:Statistics

84. **Help:Multilingual - Wikidata** https://www.wikidata.org/wiki/Help:Multilingual

85. **RDF 1.1 Semantics** https://www.w3.org/TR/rdf11-mt/

86. **Wikidata:Data access - Wikidata** https://www.wikidata.org/wiki/Wikidata:Data_access

87. **WikidataR package - RDocumentation** https://www.rdocumentation.org/packages/WikidataR/versions/2.2.0

88. **wikidata2df: Utility package for easily turning a SPARQL query into a dataframe**
João Vitor F Cavalcante
https://github.com/jvfe/wikidata2df

89. **Wikidata:Licensing - Wikidata** https://www.wikidata.org/wiki/Wikidata:Licensing

90. https://query.wikidata.org/

91. Q56010228

92. **Scholia**
Scholia
https://scholia.toolforge.org/

93. **SARS-CoV-2-Queries**
SARS-CoV-2-Queries
https://egonw.github.io/SARS-CoV-2-Queries/

94. **Wikidata:Tools/OpenRefine - Wikidata** https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine

95. **Help:QuickStatements - Wikidata** https://www.wikidata.org/wiki/Help:QuickStatements

96. **Wikidata:Bots - Wikidata** https://www.wikidata.org/wiki/Wikidata:Bots

97. **Wikidata:Pywikibot - Python 3 Tutorial - Wikidata** https://www.wikidata.org/wiki/Wikidata:Pywikibot_-_Python_3_Tutorial

98. **GitHub - SuLab/WikidataIntegrator: A Wikidata Python module integrating the MediaWiki API and the Wikidata SPARQL endpoint**
GitHub
https://github.com/SuLab/WikidataIntegrator

99. **Wikidata: A large-scale collaborative ontological medical database**
Houcemeddine Turki, Thomas Shafee, Mohamed Ali Hadj Taieb, Mohamed Ben Aouicha, Denny Vrandečić, Diptanshu Das, Helmi Hamdi
Journal of Biomedical Informatics (2019-09-23) https://www.wikidata.org/wiki/Q68471881
DOI: 10.1016/j.jbi.2019.103292

100. **Big data: Wikiomics**
Mitch Waldrop
Nature (2008-09-04) https://www.wikidata.org/wiki/Q28292893
DOI: 10.1038/455022a

101. **Calling on a million minds for community annotation in WikiProteins**
Barend Mons, Michael Ashburner, Christine Chichester, Erik M van Mulligen, Marc Weeber, Johan den Dunnen, Gert-Jan van Ommen, Mark A Musen, Matt Cockerill, Henning Hermjakob, … Amos Bairoch
Genome Biology (2008-01-01) https://www.wikidata.org/wiki/Q21183907
DOI: 10.1186/gb-2008-9-5-r89

102. **Ten Simple Rules for Developing Public Biological Databases**
Mohamed Helmy, Alexander Crits-Christoph, Gary D Bader
PLOS Computational Biology (2016-11-01) https://www.wikidata.org/wiki/Q28595967
DOI: 10.1371/journal.pcbi.1005128

103. **Inside the Alexa-Friendly World of Wikidata**
Tom Simonite
Wired https://www.wired.com/story/inside-the-alexa-friendly-world-of-wikidata/

104. **A gene wiki for community annotation of gene function**
Jon W Huss, Camilo Orozco, James Goodale, Chunlei Wu, Serge Batalov, Tim J Vickers, Faramarz Valafar, Andrew I Su
PLOS Biology (2008-07-08) https://www.wikidata.org/wiki/Q21092744
DOI: 10.1371/journal.pbio.0060175

105. **Making your database available through Wikipedia: the pros and cons**
Robert D Finn, Paul P Gardner, Alex Bateman
Nucleic Acids Research (2012-01-01) https://www.wikidata.org/wiki/Q28254676
DOI: 10.1093/nar/gkr1195

106. **Wikidata as a semantic framework for the Gene Wiki initiative**
Sebastian Burgstaller-Muehlbacher, Andra Waagmeester, Elvira Mitraka, Julia Turner, Timothy Elliott Putman, Justin Leong, Chinmay Naik, Paul Pavlidis, Lynn Schriml, Benjamin M Good, Andrew I Su
Database (2016-01-01) https://www.wikidata.org/wiki/Q23712646
DOI: 10.1093/database/baw015

107. **WikiGenomes: an open web application for community consumption and curation of gene annotation data in Wikidata**
Timothy Elliott Putman, Sebastien Lelong, Sebastian Burgstaller-Muehlbacher, Andra Waagmeester, Colin Diesh, Nathan Dunn, Monica Munoz-Torres, Gregory Stupp, Chunlei Wu, Andrew I Su, Benjamin M Good
Database (2017-03-08) https://www.wikidata.org/wiki/Q28529449

108. **ChlamBase: a curated model organism database for the Chlamydia research community**
Timothy Elliott Putman, Kevin Hybiske, Derek Jow, Cyrus Afrasiabi, Sebastien Lelong, Marco Alvarado Cano, Chunlei Wu, Andrew I Su
Database (2019-01-01) https://www.wikidata.org/wiki/Q63286185
DOI: 10.1093/database/baz041

109. **Submit a Topic Page to PLOS Computational Biology and Wikipedia**
Daniel Mietchen, Shoshana Wodak, Szymon Wasik, Natalia Szostak, Christophe Dessimoz
PLOS Computational Biology (2018-05-31) https://www.wikidata.org/wiki/Q54655231
DOI: 10.1371/journal.pcbi.1006137

110. **Scholia, Scientometrics and Wikidata**
Finn Årup Nielsen, Daniel Mietchen, Egon Willighagen
The Semantic Web: ESWC 2017 Satellite Events (2017-10-01) https://www.wikidata.org/wiki/Q41799194
DOI: 10.1007/978-3-319-70407-4_36

111. **Robustifying Scholia: paving the way for knowledge discovery and research assessment through Wikidata**
Lane Rasberry, Egon Willighagen, Finn Årup Nielsen, Daniel Mietchen
Research Ideas and Outcomes (2019-05-02) https://www.wikidata.org/wiki/Q63433973
DOI: 10.3897/rio.5.e35820

112. **Representing COVID-19 information in collaborative knowledge graphs: The case of Wikidata**
Houcemeddine Turki, Mohamed Ali Hadj Taieb, Thomas Shafee, Tiago Lubiana, Dariusz Jemielniak, Mohamed Ben Aouicha, José Emilio Labra Gayo, Eric Youngstrom, Mossab Banat, Diptanshu Das, … WikiProject COVID-19
Semantic Web: Interoperability, Usability, Applicability (2021-09-28) https://www.wikidata.org/wiki/Q108766311
DOI: 10.3233/sw-210444

113. **A protocol for adding knowledge to Wikidata: aligning resources on human coronaviruses**

Andra Waagmeester, Egon Willighagen, Andrew I Su, Martina Summer-Kutmon, José Emilio Labra Gayo, Daniel Fernández-Álvarez, Quentin J Groom, Peter J Schaap, Lisa M Verhagen, Jasper Koehorst
BMC Biology (2021-01-22) https://www.wikidata.org/wiki/Q105037759
DOI: 10.1186/s12915-020-00940-y

114. **Wikidata Queries around the SARS-CoV-2 virus and pandemic** https://www.wikidata.org/wiki/Q88647643

115. **COVIWD: COVID-19 Wikidata Dashboard**
Fariz Darari
Jurnal Ilmu Komputer dan Informasi (2021-03-01) https://www.wikidata.org/wiki/Q105833381
DOI: 10.21609/jiki.v14i1.941

116. **Painel de informação sobre a COVID-19: consultas SPARQL na Wikidata**
Ana Carolina Simionato Arakaki, Fabiano Ferreira de Castro, Felipe Augusto Arakaki
AtoZ: Novas Práticas em Informação e Conhecimento (2020-12-03) https://www.wikidata.org/wiki/Q106249454
DOI: 10.5380/atoz.v9i2.76684

117. **Uso de Wikidata y Wikipedia para la generación asistida de un vocabulario estructurado multilingüe sobre la pandemia de Covid-19**
Tomás Saorín, Juan-Antonio Pastor-Sánchez, María-José Baños-Moreno
Profesional de la Informacion (2020-09-13) https://www.wikidata.org/wiki/Q107377131
DOI: 10.3145/epi.2020.sep.09

118. **The Bgee suite: integrated curated expression atlas and comparative transcriptomics in animals**
Frederic B Bastian, Julien Roux, Anne Niknejad, Aurélie Comte, Sara SFonseca Costa, Tarcisio M Farias, Sébastien Moretti, Gilles Parmentier, Valentine Rech de Laval, Marta Rosikiewicz, ... Marc Robinson-Rechavi
Nucleic Acids Research (2020-10-10) https://www.wikidata.org/wiki/Q100513179
DOI: 10.1093/nar/gkaa793

119. **Utilizing the Wikidata system to improve the quality of medical content in Wikipedia in diverse languages: a pilot study**
Alexander Pfundner, Tobias Schönberg, John Horn, Richard David Boyce, Matthias Samwald
Journal of Medical Internet Research (2015-05-05) https://www.wikidata.org/wiki/Q21503276
DOI: 10.2196/jmir.4163

120. **Pesquisa-ação: uma introdução metodológica**
David Tripp
Educação e Pesquisa (2005-12-01) https://www.wikidata.org/wiki/Q108479295
DOI: 10.1590/s1517-97022005000300009

121. **Come si fa una tesi di laurea** https://www.wikidata.org/wiki/Q3684178

122. **The Human Cell Atlas: Technical approaches and challenges.**
Chung Chau Hon, Jay W Shin, Piero Carninci, Michael JT Stubbington
Briefings in functional genomics (2017-10-28) https://www.wikidata.org/wiki/Q48563763
DOI: 10.1093/bfgp/elx029

123. **Towards a pragmatic definition of cell type**
Tiago Lubiana, Helder I Nakaya
Authorea, Inc. (2021-01-04) https://doi.org/ghrxwf
DOI: 10.22541/au.160979530.02627436/v1

124. **PhyloCode** https://www.wikidata.org/wiki/Q1189395

References

1. **An era of single-cell genomics consortia**
Yoshinari Ando, Andrew T Kwon, Jay W Shin
Experimental and Molecular Medicine (2020-09-15) https://www.wikidata.org/wiki/Q99418649
DOI: 10.1038/s12276-020-0409-x

2. **The Human Cell Atlas.**
Aviv Regev, Sarah Teichmann, Eric Lander, Amir Giladi, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, Peter Campbell, Piero Carninci, Menna R Clatworthy, ... Human Cell Atlas Meeting Participants
eLife (2017-12-05) https://www.wikidata.org/wiki/Q46368626
DOI: 10.7554/elife.27041

3. **The Human Cell Atlas and equity: lessons learned**
Partha P Majumder, Musa M Mhlanga, Alex K Shalek
Nature Medicine (2020-10-01) https://www.wikidata.org/wiki/Q100491106
DOI: 10.1038/s41591-020-1100-4

4. **The Human Cell Atlas White Paper**
Aviv Regev, Sarah Teichmann, Orit Rozenblatt-Rosen, Michael JT Stubbington, Kristin Ardlie, Amir Giladi, Paola Arlotta, Gary D Bader, Christophe Benoist, Moshe Biton, ... Human Cell Atlas Organizing Committee
(2018-10-11) https://www.wikidata.org/wiki/Q104450645

5. **Everyone needs a data-management plan**
Nature
(2018-03-15) https://www.wikidata.org/wiki/Q56524391
DOI: 10.1038/d41586-018-03065-z

6. **About the Data Coordination Platform**
HCA Data Portal
https://data.humancellatlas.org/about/

7. **What Is Your Conceptual Definition of "Cell Type" in the Context of a Mature Organism?**
Paul Blainey, Hans Clevers, Cole Trapnell, Ed Lein, Emma Lundberg, Alfonso Martinez Arias, Joshua R Sanes, Jay Shendure, James Eberwine, Junhyong Kim, ... Mathias Uhlén
Cell systems (2017-03-01) https://www.wikidata.org/wiki/Q87649649
DOI: 10.1016/j.cels.2017.03.006

8. **A periodic table of cell types**
Bo Xia, Itai Yanai
Development (2019-06-15) https://doi.org/ggctwf
DOI: 10.1242/dev.169854 · PMID: 31249003 · PMCID: PMC6602355

9. **Exciting times to study the identity and evolution of cell types**
Maria Sachkova, Pawel Burkhardt
Development (2019-09-15) https://doi.org/ghdb9v
DOI: 10.1242/dev.178996 · PMID: 31537583

10. **The Human Cell Atlas: from vision to reality.**
Orit Rozenblatt-Rosen, Michael JT Stubbington, Aviv Regev, Sarah Teichmann
Nature (2017-10-01) https://www.wikidata.org/wiki/Q47565008
DOI: 10.1038/550451a

11. **Human Cell Atlas and cell-type authentication for regenerative medicine**
Yulia Panina, Peter Karagiannis, Andreas Kurtz, Glyn N Stacey, Wataru Fujibuchi
Experimental and Molecular Medicine (2020-09-15) https://www.wikidata.org/wiki/Q99418657
DOI: 10.1038/s12276-020-0421-1

12. **A community-based transcriptomics classification and nomenclature of neocortical cell types**
Rafael Yuste, Michael J Hawrylycz, Nadia Aalling, Argel Aguilar-Valles, Detlev Arendt, Rubén Armañanzas, Giorgio A Ascoli, Concha Bielza, Vahid Bokharaie, Tobias B Bergmann, ... Ed S Lein
Nature Neuroscience (2020-08-24) https://www.wikidata.org/wiki/Q98665291
DOI: 10.1038/s41593-020-0685-8

13. **The evolving concept of cell identity in the single cell era**
Samantha A Morris
Development (2019-06-27) https://www.wikidata.org/wiki/Q93086971
DOI: 10.1242/dev.169748

14. **Implications of Epigenetic Variability within a Cell Population for "Cell Type" Classification**
Inna Tabansky, Joel Stern, Donald W Pfaff
Frontiers in Behavioral Neuroscience (2015-12-16) https://www.wikidata.org/wiki/Q26770736
DOI: 10.3389/fnbeh.2015.00342

15. **Geometry of the Gene Expression Space of Individual Cells**
Yael Korem, Pablo Szekely, Yuval Hart, Hila Sheftel, Jean Hausser, Avi Mayo, Michael E Rothenberg, Tomer Kalisky, Uri Alon
PLOS Computational Biology (2015-07-10) https://www.wikidata.org/wiki/Q35688096
DOI: 10.1371/journal.pcbi.1004224

16. **Evolution of Cellular Differentiation: From Hypotheses to Models**
Pedro Márquez-Zacarías, Rozenn M Pineau, Marcella Gomez, Alan Veliz-Cuba, David Murrugarra, William C Ratcliff, Karl J Niklas
Trends in Ecology & Evolution (2020-08-20) https://www.wikidata.org/wiki/Q98633613

DOI: 10.1016/j.tree.2020.07.013

17. **Inferring cell type innovations by phylogenetic methods-concepts, methods, and limitations**
Koryu Kin, Koryu Kin
Journal of Experimental Zoology. Part B: Molecular and Developmental Evolution (2015-10-14) https://www.wikidata.org/wiki/Q40436539
DOI: 10.1002/jez.b.22657

18. **Towards a pragmatic definition of cell type**
Tiago Lubiana, Helder Nakaya
(2021-01-04) https://www.wikidata.org/wiki/Q108723646
DOI: 10.22541/au.160979530.02627436/v1

19. **BioNumbers--the database of key numbers in molecular and cell biology**
Ron Milo, Paul Jorgensen, Uri Moran, Griffin M Weber, Michael Springer
Nucleic Acids Research (2010-01-01) https://www.wikidata.org/wiki/Q24643881
DOI: 10.1093/nar/gkp889

20. **Wikipedia, the free encyclopedia** https://en.wikipedia.org/wiki/Main_Page

21. **Cell types and ontologies of the Human Cell Atlas**
David Osumi-Sutherland, Chuan Xu, Maria C Keays, Peter V Kharchenko, Aviv Regev, Ed S Lein, Sarah Teichmann
(2021-06-28) https://www.wikidata.org/wiki/Q107373831

22. **An estimation of the number of cells in the human body.**
Eva Bianconi, Allison Piovesan, Federica Facchin, Alina Beraudi, Raffaella Casadei, Flavia Frabetti, Lorenza Vitale, Maria Chiara Pelleri, Simone Tassani, Francesco Piva, … Silvia Canaider
Annals of Human Biology (2013-07-05) https://www.wikidata.org/wiki/Q34037445
DOI: 10.3109/03014460.2013.807878

23. **A curated database reveals trends in single-cell transcriptomics**
Valentine Svensson, Eduardo da Veiga Beltrame, Lior Pachter
Database (2020-11-01) https://www.wikidata.org/wiki/Q103034964
DOI: 10.1093/database/baaa073

24. **The evolution of cell types in animals: emerging principles from molecular studies.**
Detlev Arendt
Nature reviews. Genetics (2008-11) https://www.ncbi.nlm.nih.gov/pubmed/18927580
DOI: 10.1038/nrg2416 · PMID: 18927580

25. **The origin and evolution of cell types**
Detlev Arendt, Jacob M Musser, Clare VH Baker, Aviv Bergman, Connie Cepko, Douglas H Erwin, Mihaela Pavlicev, Gerhard Schlosser, Stefanie Widder, Manfred D Laubichler, Günter P Wagner
Nature Reviews Genetics (2016-11-07) https://doi.org/f9b62x
DOI: 10.1038/nrg.2016.127 · PMID: 27818507

26. **Stem cell states, fates, and the rules of attraction.**
Tariq Enver, Martin Pera, Carsten Peterson, Peter W Andrews
Cell Stem Cell (2009-05-01) https://www.wikidata.org/wiki/Q37475461
DOI: 10.1016/j.stem.2009.04.011

27. **Theory of cell fate**
Michael J Casey, Patrick S Stumpf, Ben D MacArthur
Wiley interdisciplinary reviews. Systems biology and medicine (2019-12-12) https://www.wikidata.org/wiki/Q91908361
DOI: 10.1002/wsbm.1471

28. **Perspectives on defining cell types in the brain**
Eran A Mukamel, John Ngai
Current Opinion in Neurobiology (2018-12-06) https://www.wikidata.org/wiki/Q90361677
DOI: 10.1016/j.conb.2018.11.007

29. **Ensembles, dynamics, and cell types: Revisiting the statistical mechanics perspective on cellular regulation**
Stefan Bornholdt, Stuart Kauffman
Journal of Theoretical Biology (2019-01-31) https://www.wikidata.org/wiki/Q91316993
DOI: 10.1016/j.jtbi.2019.01.036

30. **Species Concepts and Species Delimitation**
Kevin De Queiroz
Systematic Biology (2007-12) https://doi.org/c34kzf
DOI: 10.1080/10635150701701083 · PMID: 18027281

31. **An ontology for cell types**
Jonathan Bard, Sue Rhee, Michael Ashburner
Genome Biology (2005-01-01) https://www.wikidata.org/wiki/Q21184168
DOI: 10.1186/gb-2005-6-2-r21

32. **Logical Development of the Cell Ontology**
Terrence F Meehan, Anna Maria Masci, Amina Abdulla, Lindsay G Cowell, Judith A Blake, Christopher J Mungall, Alexander D Diehl
BMC Bioinformatics (2011-01-05) https://doi.org/c7kw6x
DOI: 10.1186/1471-2105-12-6 · PMID: 21208450 · PMCID: PMC3024222

33. **The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability**
Alexander D Diehl, Terrence F Meehan, Yvonne M Bradford, Matthew H Brush, Wasila M Dahdul, David S Dougall, Yongqun He, David Osumi-Sutherland, Alan Ruttenberg, Sirarat Sarntivijai, … Christopher J Mungall
Journal of Biomedical Semantics (2016-07-04) https://doi.org/gg99b9

DOI: 10.1186/s13326-016-0088-7 · PMID: 27377652 · PMCID: PMC4932724

34. **CELDA -- an ontology for the comprehensive representation of cells in complex systems**
Stefanie Seltmann, Harald Stachelscheid, Alexander Damaschun, Ludger Jansen, Fritz Lekschas, Jean-Fred Fontaine, Throng Nghia Nguyen-Dobinsky, Ulf Leser, Andreas Kurtz
BMC Bioinformatics (2013-07-17) https://www.wikidata.org/wiki/Q21284308
DOI: 10.1186/1471-2105-14-228

35. **Cells in experimental life sciences - challenges and solution to the rapid evolution of knowledge**
Sirarat Sarntivijai, Alexander D Diehl, Yongqun He
BMC Bioinformatics (2017-12-21) https://doi.org/gg99b7
DOI: 10.1186/s12859-017-1976-2 · PMID: 29322916 · PMCID: PMC5763506

36. **Cells in ExperimentaL Life Sciences (CELLS-2018): capturing the knowledge of normal and diseased cells with ontologies**
Sirarat Sarntivijai, Yongqun He, Alexander D Diehl
BMC Bioinformatics (2019-04-25) https://doi.org/gg99b8
DOI: 10.1186/s12859-019-2721-9 · PMID: 31272374 · PMCID: PMC6509796

37. **Scaled, high fidelity electrophysiological, morphological, and transcriptomic cell characterization**
Brian R Lee, Agata Budzillo, Kristen Hadley, Jeremy A Miller, Tim Jarsky, Katherine Baker, DiJon Hill, Lisa Kim, Rusty Mann, Lindsay Ng, … Jim Berg
eLife (2021-08-13) https://www.wikidata.org/wiki/Q109717199
DOI: 10.7554/elife.65482

38. **Current best practices in single-cell RNA-seq analysis: a tutorial**
Malte D Luecken, Fabian J Theis
Molecular Systems Biology (2019-06-19) https://www.wikidata.org/wiki/Q64974172
DOI: 10.15252/msb.20188746

39. **Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods**
Zoe A Clarke, Tallulah Andrews, Jawairia Atif, Delaram Pouyabahar, Brendan T Innes, Sonya A MacParland, Gary D Bader
Nature Protocols (2021-05-24) https://www.wikidata.org/wiki/Q107158224
DOI: 10.1038/s41596-021-00534-0

40. **Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis**
Jacob H Levine, Erin F Simonds, Sean C Bendall, Kara L Davis, El-ad D Amir, Michelle D Tadmor, Oren Litvin, Harris G Fienberg, Astraea Jager, Eli R Zunder, … Garry P Nolan
Cell (2015-06-18) https://www.wikidata.org/wiki/Q30975629
DOI: 10.1016/j.cell.2015.05.047

41. **Fast unfolding of communities in large networks**
Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre
Journal of Statistical Mechanics: Theory and Experiment (2008-10-09) https://www.wikidata.org/wiki/Q29305711
DOI: 10.1088/1742-5468/2008/10/p10008

42. **PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data**
Oscar Franzén, Li-Ming Gan, Johan LM Bjorkegren
Database (2019-01-01) https://www.wikidata.org/wiki/Q63664483
DOI: 10.1093/database/baz046

43. **CellMarker: a manually curated resource of cell markers in human and mouse**
Xinxin Zhang, Yujia Lan, Jinyuan Xu, Fei Quan, Erjie Zhao, Chunyu Deng, Tao Luo, Liwen Xu, Gaoming Liao, Min Yan, … Yun Xiao
Nucleic Acids Research (2019-01-01) https://www.wikidata.org/wiki/Q56984510
DOI: 10.1093/nar/gky900

44. **Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data**
Julie A McMurry, Nick Juty, Niklas Blomberg, Tony Burdett, Tom Conlin, Nathalie Conte, Melanie Courtot, John Deck, Michel Dumontier, Donal K Fellows, … Helen Parkinson
PLOS Biology (2017-06-29) https://www.wikidata.org/wiki/Q33037209
DOI: 10.1371/journal.pbio.2001414

45. **The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability.**
Alexander D Diehl, Terrence F Meehan, Yvonne M Bradford, Matthew H Brush, Wasila M Dahdul, David S Dougall, Yongqun He, David Osumi-Sutherland, Alan Ruttenberg, Sirarat Sarntivijai, … Christopher J Mungall
Journal of Biomedical Semantics (2016-07-04) https://www.wikidata.org/wiki/Q36067763
DOI: 10.1186/s13326-016-0088-7

46. **Cell type discovery using single-cell transcriptomics: implications for ontological representation**
Brian D Aevermann, Mark Novotny, Trygve E Bakken, Jeremy A Miller, Alexander D Diehl, David Osumi-Sutherland, Roger S Lasken, Ed S Lein, Richard H Scheuermann
Human Molecular Genetics (2018-05-01) https://www.wikidata.org/wiki/Q52625486
DOI: 10.1093/hmg/ddy100

47. **Cell ontology in an age of data-driven cell classification.**
David Osumi-Sutherland, David Osumi-Sutherland
BMC Bioinformatics (2017-12-21) https://www.wikidata.org/wiki/Q49192555
DOI: 10.1186/s12859-017-1980-6

48. **Cell type ontologies of the Human Cell Atlas**
David Osumi-Sutherland, Chuan Xu, Maria Keays, Adam P Levine, Peter V Kharchenko, Aviv Regev, Ed Lein, Sarah Teichmann
Nature Cell Biology (2021-11-01) https://www.wikidata.org/wiki/Q109755180
DOI: 10.1038/s41556-021-00787-7

49. **Besca, a single-cell transcriptomics analysis toolkit to accelerate translational research**

Sophia Clara Mädler, Alice Julien-Laferriere, Luis Wyss, Miroslav Phan, Albert SW Kang, Eric Ulrich, Roland Schmucki, Jitao David Zhang, Martin Ebeling, Laura Badi, … Klas Hatje
bioRxiv (2020-08-12) https://www.wikidata.org/wiki/Q104450593
DOI: 10.1101/2020.08.11.245795

50. **Leveraging the Cell Ontology to classify unseen cell types**
Sheng Wang, Angela Oliveira Pisco, Aaron McGeever, Maria Brbić, Marinka Žitnik, Spyros Darmanis, Jure Leskovec, Jim Karkanias, Russ Altman
Nature Communications (2021-09-21) https://www.wikidata.org/wiki/Q108929315
DOI: 10.1038/s41467-021-25725-x

51. **ontoProc: processing of ontologies of anatomy, cell lines, and so on** https://www.wikidata.org/wiki/Q101074371

52. **Tabula Muris** https://tabula-muris.ds.czbiohub.org/

53. **Tabula Sapiens** https://tabula-sapiens-portal.ds.czbiohub.org/celltypes

54. **Azimuth** https://azimuth.hubmapconsortium.org/

55. **Construction and Usage of a Human Body Common Coordinate Framework Comprising Clinical, Semantic, and Spatial Ontologies**
Katy Börner, Ellen Quardokus, Bruce WHerr II, Leonard E Cross, Elizabeth G Record, Yingnan Ju, Andreas D Bueckle, James P Sluka, Jonathan C Silverstein, Kristen M Browne, … Griffin M Weber
(2020-07-28) https://www.wikidata.org/wiki/Q109755184

56. **Cell Annotation Platform | Coming Soon** http://celltype.info/

57. **Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture across the human body**
Conde C Domínguez, Tomás Gomes, Lorna B Jarvis, C Xu, SK Howlett, DB Rainbow, Ondrej Suchanek, Hamish W King, Lira Mamanova, Krzysztof Polański, … Sarah Teichmann
(2021-04-28) https://www.wikidata.org/wiki/Q107363182
DOI: 10.1101/2021.04.28.441762

58. **Ontology based molecular signatures for immune cell types via gene expression analysis**
Terrence F Meehan, Nicole Vasilevsky, Christopher J Mungall, David S Dougall, Melissa Haendel, Judith A Blake, Alexander D Diehl
BMC Bioinformatics (2013-08-30) https://www.wikidata.org/wiki/Q34978215
DOI: 10.1186/1471-2105-14-263

59. **Logical development of the cell ontology**
Terrence F Meehan, Anna Maria Masci, Amina Abdulla, Lindsay G Cowell, Judith A Blake, Christopher J Mungall, Alexander D Diehl
BMC Bioinformatics (2011-01-05) https://www.wikidata.org/wiki/Q33786317
DOI: 10.1186/1471-2105-12-6

60. **Scientists are drowning in COVID-19 papers. Can new tools keep them afloat?** https://www.science.org/content/article/scientists-are-drowning-covid-19-papers-can-new-tools-keep-them-afloat

61. **How a torrent of COVID science changed research publishing — in seven charts**
Holly Else
Nature (2020-12-16) https://www.nature.com/articles/d41586-020-03564-y

62. **5-star Open Data** http://5stardata.info/en/

63. **UniProt: the universal protein knowledgebase in 2021**
UniProt Consortium
Nucleic Acids Research (2020-11-25) https://www.wikidata.org/wiki/Q102383737
DOI: 10.1093/nar/gkaa1100

64. **The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets**
Damian Szklarczyk, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, Sampo Pyysalo, Nadezhda Tsankova Doncheva, Marc Legeay, Tao Fang, Peer Bork, … Christian von Mering
Nucleic Acids Research (2020-11-25) https://www.wikidata.org/wiki/Q102383784
DOI: 10.1093/nar/gkaa1074

65. **Wikidata as a knowledge graph for the life sciences**
Andra Waagmeester, Gregory Stupp, Sebastian Burgstaller-Muehlbacher, Benjamin M Good, Malachi Griffith, Obi Griffith, Kristina Hanspers, Henning Hermjakob, Toby Hudson, Kevin Hybiske, … Andrew I Su
eLife (2020-03-17) https://www.wikidata.org/wiki/Q87830400
DOI: 10.7554/elife.52614

66. **Literature-based discovery - Wikipedia** https://en.wikipedia.org/wiki/Literature-based_discovery

67. **Literature Based Discovery: models, methods, and trends.**
MSSam Henry, Bridget T McInnes
Journal of Biomedical Informatics (2017-08-21) https://www.wikidata.org/wiki/Q38371706
DOI: 10.1016/j.jbi.2017.08.011

68. **Online tools to support literature-based discovery in the life sciences.**
Marc Weeber, Marc Weeber, Jan A Kors, Jan A Kors, Barend Mons
Briefings in Bioinformatics (2005-09-01) https://www.wikidata.org/wiki/Q36280460
DOI: 10.1093/bib/6.3.277

69. **Unsupervised word embeddings capture latent knowledge from materials science literature**
Vahe Tshitoyan, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A Persson, Gerbrand Ceder, Anubhav Jain
Nature (2019-07-03) https://www.wikidata.org/wiki/Q91595456
DOI: 10.1038/s41586-019-1335-8

70. **Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis.**
Nadine Bakkar, Tina Kovalik, Ileana Lorenzini, Scott Spangler, Alix Lacoste, Kyle Sponaugle, Philip Ferrante, Elenee Argentinis, Rita Sattler, Robert Bowser
Acta Neuropathologica (2017-11-13) https://www.wikidata.org/wiki/Q47406275
DOI: 10.1007/s00401-017-1785-8

71. **Ontologies for the life sciences**
Steffen Schulze-Kremer, Barry Smith
(2005-11-15) https://www.wikidata.org/wiki/Q105870680
DOI: 10.1002/047001153x.g408213

72. **The Philosophy of Logical Atomism, Lecture 1: Facts and Propositions** https://www.wikidata.org/wiki/Q105105637

73. **Logik der Forschung**
Karl Popper
(1934-01-01) https://www.wikidata.org/wiki/Q1868040

74. **The semantic conception of truth: and the foundations of semantics**
Alfred Tarski
Philosophy and Phenomenological Research (1944-03-01) https://www.wikidata.org/wiki/Q106090790
DOI: 10.2307/2102968

75. **The Gene Ontology resource: enriching a GOld mine**
Gene Ontology Consortium
Nucleic Acids Research (2020-12-08) https://www.wikidata.org/wiki/Q104130127
DOI: 10.1093/nar/gkaa1113

76. **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**
M Ashburner, CA Ball, Judith A Blake, David Botstein, H Butler, JMichael Cherry, AP Davis, K Dolinski, Selina S Dwight, JT Eppig, … Gavin Sherlock
Nature Genetics (2000-05-01) https://www.wikidata.org/wiki/Q23781406
DOI: 10.1038/75556

77. **The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration**
Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J Mungall, … Suzanna Lewis
Nature Biotechnology (2007-11-01) https://www.wikidata.org/wiki/Q19671692
DOI: 10.1038/nbt1346

78. **Introducing the Knowledge Graph: things, not strings**
Google
(2012-05-16) https://blog.google/products/search/introducing-knowledge-graph-things-not/

79. **Toward an epistemology of Wikipedia**
Don Fallis
Journal of the Association for Information Science and Technology (2008-08-01) https://www.wikidata.org/wiki/Q101955295
DOI: 10.1002/asi.20870

80. **From Freebase to Wikidata: The Great Migration**
Thomas Pellissier Tanon, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, Lydia Pintscher
Proceedings of the 25th International Conference on World Wide Web (2016-01-01) https://www.wikidata.org/wiki/Q24074986
DOI: 10.1145/2872427.2874809

81. **Wikibase/DataModel - MediaWiki** https://www.mediawiki.org/wiki/Wikibase/DataModel

82. **Help:Data type - Wikidata** https://www.wikidata.org/wiki/Help:Data_type

83. **Wikidata:Statistics - Wikidata** https://www.wikidata.org/wiki/Wikidata:Statistics

84. **Help:Multilingual - Wikidata** https://www.wikidata.org/wiki/Help:Multilingual

85. **RDF 1.1 Semantics** https://www.w3.org/TR/rdf11-mt/

86. **Wikidata:Data access - Wikidata** https://www.wikidata.org/wiki/Wikidata:Data_access

87. **WikidataR package - RDocumentation** https://www.rdocumentation.org/packages/WikidataR/versions/2.2.0

88. **wikidata2df: Utility package for easily turning a SPARQL query into a dataframe**
João Vitor F Cavalcante
https://github.com/jvfe/wikidata2df

89. **Wikidata:Licensing - Wikidata** https://www.wikidata.org/wiki/Wikidata:Licensing

90. https://query.wikidata.org/

91. Q56010228

92. **Scholia**
Scholia
https://scholia.toolforge.org/

93. **SARS-CoV-2-Queries**
SARS-CoV-2-Queries
https://egonw.github.io/SARS-CoV-2-Queries/

94. **Wikidata:Tools/OpenRefine - Wikidata** https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine

95. **Help:QuickStatements - Wikidata** https://www.wikidata.org/wiki/Help:QuickStatements

96. **Wikidata:Bots - Wikidata** https://www.wikidata.org/wiki/Wikidata:Bots

97. **Wikidata:Pywikibot - Python 3 Tutorial - Wikidata** https://www.wikidata.org/wiki/Wikidata:Pywikibot_-_Python_3_Tutorial

98. **GitHub - SuLab/WikidataIntegrator: A Wikidata Python module integrating the MediaWiki API and the Wikidata SPARQL endpoint**
GitHub
https://github.com/SuLab/WikidataIntegrator

99. **Wikidata: A large-scale collaborative ontological medical database**
Houcemeddine Turki, Thomas Shafee, Mohamed Ali Hadj Taieb, Mohamed Ben Aouicha, Denny Vrandečić, Diptanshu Das, Helmi Hamdi
Journal of Biomedical Informatics (2019-09-23) https://www.wikidata.org/wiki/Q68471881
DOI: 10.1016/j.jbi.2019.103292

100. **Big data: Wikiomics**
Mitch Waldrop
Nature (2008-09-04) https://www.wikidata.org/wiki/Q28292893
DOI: 10.1038/455022a

101. **Calling on a million minds for community annotation in WikiProteins**
Barend Mons, Michael Ashburner, Christine Chichester, Erik M van Mulligen, Marc Weeber, Johan den Dunnen, Gert-Jan van Ommen, Mark A Musen, Matt Cockerill, Henning Hermjakob, … Amos Bairoch
Genome Biology (2008-01-01) https://www.wikidata.org/wiki/Q21183907
DOI: 10.1186/gb-2008-9-5-r89

102. **Ten Simple Rules for Developing Public Biological Databases**
Mohamed Helmy, Alexander Crits-Christoph, Gary D Bader
PLOS Computational Biology (2016-11-01) https://www.wikidata.org/wiki/Q28595967
DOI: 10.1371/journal.pcbi.1005128

103. **Inside the Alexa-Friendly World of Wikidata**
Tom Simonite
Wired https://www.wired.com/story/inside-the-alexa-friendly-world-of-wikidata/

104. **A gene wiki for community annotation of gene function**
Jon W Huss, Camilo Orozco, James Goodale, Chunlei Wu, Serge Batalov, Tim J Vickers, Faramarz Valafar, Andrew I Su
PLOS Biology (2008-07-08) https://www.wikidata.org/wiki/Q21092744
DOI: 10.1371/journal.pbio.0060175

105. **Making your database available through Wikipedia: the pros and cons**
Robert D Finn, Paul P Gardner, Alex Bateman
Nucleic Acids Research (2012-01-01) https://www.wikidata.org/wiki/Q28254676
DOI: 10.1093/nar/gkr1195

106. **Wikidata as a semantic framework for the Gene Wiki initiative**
Sebastian Burgstaller-Muehlbacher, Andra Waagmeester, Elvira Mitraka, Julia Turner, Timothy Elliott Putman, Justin Leong, Chinmay Naik, Paul Pavlidis, Lynn Schriml, Benjamin M Good, Andrew I Su
Database (2016-01-01) https://www.wikidata.org/wiki/Q23712646
DOI: 10.1093/database/baw015

107. **WikiGenomes: an open web application for community consumption and curation of gene annotation data in Wikidata**
Timothy Elliott Putman, Sebastien Lelong, Sebastian Burgstaller-Muehlbacher, Andra Waagmeester, Colin Diesh, Nathan Dunn, Monica Munoz-Torres, Gregory Stupp, Chunlei Wu, Andrew I Su, Benjamin M Good
Database (2017-03-08) https://www.wikidata.org/wiki/Q28529449

108. **ChlamBase: a curated model organism database for the Chlamydia research community**
Timothy Elliott Putman, Kevin Hybiske, Derek Jow, Cyrus Afrasiabi, Sebastien Lelong, Marco Alvarado Cano, Chunlei Wu, Andrew I Su
Database (2019-01-01) https://www.wikidata.org/wiki/Q63286185
DOI: 10.1093/database/baz041

109. **Submit a Topic Page to PLOS Computational Biology and Wikipedia**
Daniel Mietchen, Shoshana Wodak, Szymon Wasik, Natalia Szostak, Christophe Dessimoz
PLOS Computational Biology (2018-05-31) https://www.wikidata.org/wiki/Q54655231
DOI: 10.1371/journal.pcbi.1006137

110. **Scholia, Scientometrics and Wikidata**
Finn Årup Nielsen, Daniel Mietchen, Egon Willighagen
The Semantic Web: ESWC 2017 Satellite Events (2017-10-01) https://www.wikidata.org/wiki/Q41799194
DOI: 10.1007/978-3-319-70407-4_36

111. **Robustifying Scholia: paving the way for knowledge discovery and research assessment through Wikidata**
Lane Rasberry, Egon Willighagen, Finn Årup Nielsen, Daniel Mietchen
Research Ideas and Outcomes (2019-05-02) https://www.wikidata.org/wiki/Q63433973
DOI: 10.3897/rio.5.e35820

112. **Representing COVID-19 information in collaborative knowledge graphs: The case of Wikidata**
Houcemeddine Turki, Mohamed Ali Hadj Taieb, Thomas Shafee, Tiago Lubiana, Dariusz Jemielniak, Mohamed Ben Aouicha, José Emilio Labra Gayo, Eric Youngstrom, Mossab Banat, Diptanshu Das, … WikiProject COVID-19
Semantic Web: Interoperability, Usability, Applicability (2021-09-28) https://www.wikidata.org/wiki/Q108766311
DOI: 10.3233/sw-210444

113. **A protocol for adding knowledge to Wikidata: aligning resources on human coronaviruses**

Andra Waagmeester, Egon Willighagen, Andrew I Su, Martina Summer-Kutmon, José Emilio Labra Gayo, Daniel Fernández-Álvarez, Quentin J Groom, Peter J Schaap, Lisa M Verhagen, Jasper Koehorst
BMC Biology (2021-01-22) https://www.wikidata.org/wiki/Q105037759
DOI: 10.1186/s12915-020-00940-y

114. **Wikidata Queries around the SARS-CoV-2 virus and pandemic** https://www.wikidata.org/wiki/Q88647643

115. **COVIWD: COVID-19 Wikidata Dashboard**
Fariz Darari
Jurnal Ilmu Komputer dan Informasi (2021-03-01) https://www.wikidata.org/wiki/Q105833381
DOI: 10.21609/jiki.v14i1.941

116. **Painel de informação sobre a COVID-19: consultas SPARQL na Wikidata**
Ana Carolina Simionato Arakaki, Fabiano Ferreira de Castro, Felipe Augusto Arakaki
AtoZ: Novas Práticas em Informação e Conhecimento (2020-12-03) https://www.wikidata.org/wiki/Q106249454
DOI: 10.5380/atoz.v9i2.76684

117. **Uso de Wikidata y Wikipedia para la generación asistida de un vocabulario estructurado multilingüe sobre la pandemia de Covid-19**
Tomás Saorín, Juan-Antonio Pastor-Sánchez, María-José Baños-Moreno
Profesional de la Informacion (2020-09-13) https://www.wikidata.org/wiki/Q107377131
DOI: 10.3145/epi.2020.sep.09

118. **The Bgee suite: integrated curated expression atlas and comparative transcriptomics in animals**
Frederic B Bastian, Julien Roux, Anne Niknejad, Aurélie Comte, Sara SFonseca Costa, Tarcisio M Farias, Sébastien Moretti, Gilles Parmentier, Valentine Rech de Laval, Marta Rosikiewicz, ... Marc Robinson-Rechavi
Nucleic Acids Research (2020-10-10) https://www.wikidata.org/wiki/Q100513179
DOI: 10.1093/nar/gkaa793

119. **Utilizing the Wikidata system to improve the quality of medical content in Wikipedia in diverse languages: a pilot study**
Alexander Pfundner, Tobias Schönberg, John Horn, Richard David Boyce, Matthias Samwald
Journal of Medical Internet Research (2015-05-05) https://www.wikidata.org/wiki/Q21503276
DOI: 10.2196/jmir.4163

120. **Pesquisa-ação: uma introdução metodológica**
David Tripp
Educação e Pesquisa (2005-12-01) https://www.wikidata.org/wiki/Q108479295
DOI: 10.1590/s1517-97022005000300009

121. **Come si fa una tesi di laurea** https://www.wikidata.org/wiki/Q3684178

122. **The Human Cell Atlas: Technical approaches and challenges.**
Chung Chau Hon, Jay W Shin, Piero Carninci, Michael JT Stubbington
Briefings in functional genomics (2017-10-28) https://www.wikidata.org/wiki/Q48563763
DOI: 10.1093/bfgp/elx029

123. **Towards a pragmatic definition of cell type**
Tiago Lubiana, Helder I Nakaya
Authorea, Inc. (2021-01-04) https://doi.org/ghrxwf
DOI: 10.22541/au.160979530.02627436/v1

124. **PhyloCode** https://www.wikidata.org/wiki/Q1189395