语义单元：将知识图谱组织成语义上有意义的表示单位

Vogt, Lars[1]; Kuhn, Tobias[2]; Hoehndorf, Robert[3]

[1] *TIB* 莱布尼茨科学与技术信息中心，*Welfengarten 1B*，*30167* 汉诺威，德国，orcid.org/0000-0002-8280-0487

[2] 计算机科学系，阿姆斯特丹自由大学，荷兰，orcid.org/0000-0002-1267-0234

[3] 计算生物科学研究中心，计算机、电气与数学科学及工程学部，阿卜杜拉国王科技大学，*KAUST 4700*，*23955* 图瓦勒，沙特阿拉伯，orcid.org/0000-0001-8149-5890

通信地址：lars.m.vogt@googlemail.com

摘要

背景：知识图谱和本体正日益成为实现可查找、可访问、可互操作和可重用的数据与元数据（FAIR指导原则）的重要技术解决方案。我们讨论了阻碍FAIR知识图谱应用的四项挑战。

结果：语义单元有望通过将知识图谱结构化为可识别且具有语义意义的子图来解决这些挑战。每个语义单元都由其自身的资源表示，实例化相应的语义单元类，并且可以实现为 FAIR数字对象和 RDF/OWL 与属性图中的纳米出版物。我们将陈述单元和复合单元区分为语义单元的基本类别。陈述单元表示对人类读者具有语义意义的最小独立命题。它们由一个或多个三元组组成，并在数学上划分知识图谱。我们将断言式、偶然式（原型式）和全称陈述单元区分为陈述单元的基本类型，并为其提出表征方案和形式语义（包括缺失陈述、否定和基数限制），这些方案和语义不涉及空白节点，且可转换回 OWL。另一方面，复合单元表示具有语义意义的语义单元集合；我们区分了多种复合单元类型，它们分别表示不同层次的表征粒度、不同类型的粒度树以及不同的参照框架。

结论：语义单元支持关于陈述的陈述，可用于图对齐、子图匹配、知识图谱剖析，以及管理敏感数据的访问限制。将图组织成语义单元，有助于分离本体信息、诊断信息（即指称信息）和话语信息，也有助于区分多个参照框架。

关键词：FAIR数据与元数据、知识图谱、OWL、RDF、语义单元、断言式陈述、偶然陈述、原型陈述、全称陈述、否定

背景

在每天生成的数据量不断增加的时代（1–3），科学和工业领域面临着新的技术与社会挑战（4），需要以创新方式表征和管理数据。能够收集、整合和分析来自各种源的大量数据，也是应对生物多样性丧失和气候变化这两大全球重大挑战的必要条件之一（5）。解决这些问题有赖于在众多利益相关方之间共享数据，因此需要努力促进来自不同领域的数据提供者相互关联，这往往要求采用真正的跨学科方法（6）。在这种方法中，数据管理仍由领域专家或机构掌握，从而确保他们拥有技术自主权（遵循 Barend Mons 的数据访问理念，而非数据共享（7））。

从数据管理和数据表示的角度来看，这要求数据和元数据符合**FAIR**原则，即能够轻松**F**查找、**A**访问、**I**互操作和**R**重复使用，适用于**机器和人类**（8）。如果做不到这一点，大数据最终就会变成暗数据（9）。将 FAIR 指导原则确立为科学和工业领域的通用标准，也有助于解决科学领域的可重复性危机（10），以及普遍存在的信息可信度问题（另见**TRUST**原则：透明度（**T**ransparency）、责任（**R**esponsibility）、用户关注（**U**ser Focus）、可持续性（**S**ustainability）和技术（**T**echnology）（11））。因此，我们必须建设类似 FAIR数据与服务互联网（12）的体系，使其能够随大数据规模扩展；所有相关的数据丰富型机构、研究项目和公民科学项目都可以据此遵循 FAIR 指导原则，使其数据和元数据可访问（13,14）。这需要提供丰富的、机器可操作的数据和元数据，同时提供人类可读的界面输出和搜索功能，并将这些数据组织为**FAIR数字对象**（15,16），每个对象都有自己的唯一持久可解析标识符（UPRI），以便单独引用。

在此背景下，**知识图谱**能够为所需的技术解决方案作出重要贡献，为管理和表示 FAIR 数据与元数据提供合适的框架（17）。知识图谱正日益普及（18），尤其是谷歌于 2012 年宣布推出 Google 知识图谱（19）之后；此后，工业界陆续宣布正在开发知识图谱，关于知识图谱的科学出版物数量也不断增长（20）。除了在工业和研究领域的一般应用外，知识图谱还特别用于基于实体和关系的语义搜索、深度推理、自然语言消歧、机器阅读，以及面向大数据和文本分析的实体整合（21）。

与关系模型或其他 NoSQL 模型相比，知识图谱采用的基于图的抽象具有多项优势，包括：（i）以直观的方式建模关系；（ii）允许推迟对数据模式定义的具体规定，使其能够灵活演进，这在处理不完整知识时尤为重要；（iii）采用本体和规则等机器可操作的知识表示形式化体系；（iv）应用图分析和机器学习；以及（v）利用知识图谱专用的图查询语言。除了支持连接、并集和投影等标准关系运算符外，这些查询语言还支持导航运算符，可沿任意长度的路径递归搜索实体（20,22–27）。此外，由于知识图谱具有固有的语义透明性，它们能够提升基于数据的决策的透明度，并改善研究和科学领域的整体沟通。

然而，尽管知识图谱为记录数据和元数据提供了合适的技术框架，但使用知识图谱并不一定会带来 FAIR 数据与元数据；要实现这一点，还需遵循特定指南，例如始终采用适当的语义数据模型，并将数据组织成 FAIR数字对象（15,16）。此外，与新技术常见的情况一样，知识图谱也带来了自身特有的技术、概念和社会挑战。首先，知识图谱这一概念本身就比较模糊（20），涵盖了不同的技术和概念形态，包括 Neo4J 等属性图（https://neo4j.com/），以及基于资源描述框架（RDF）的方法、使用RDF 存储库的做法，以及通过 Web 本体语言（OWL）实现的描述逻辑应用。

在此，我们先简要讨论其中四项挑战，随后介绍将知识图谱划分并组织为**可识别且具有语义意义的表征单元（简称：语义单元）**这一构想。语义单元的概念能够为解决这四项挑战作出重要贡献。我们将语义单元分为两种基本类别，即陈述单元和复合单元，并将它们作为 FAIR 知识图谱中的新元素，与广为人知的三元组及整体图并列。陈述单元和复合单元可用于将数据图组织为五个表征粒度层级，范围从单个三元组层级到整体图层级。我们还介绍了语义单元的其他子类别，可用于进一步组织数据图。我们接着论证，由于每个语义单元都可以组织为具有自身 UPRI 的 FAIR数字对象，因此可以在三元组陈述中引用语义单元，从而以非常高效的方式提出关于陈述的陈述。引入语义单元后，我们采用以用户为中心的方法，并**在成熟的知识图谱 RDF 和 OWL 层之上再增加一层三元组**（图 1）。通过简化经验数据的语义建模并降低其

表征复杂性，并为那些在 OWL 中不存在形式语义的陈述提供表征（即模式）和形式语义，语义单元提高了知识图谱对领域专家和开发者的可用性。

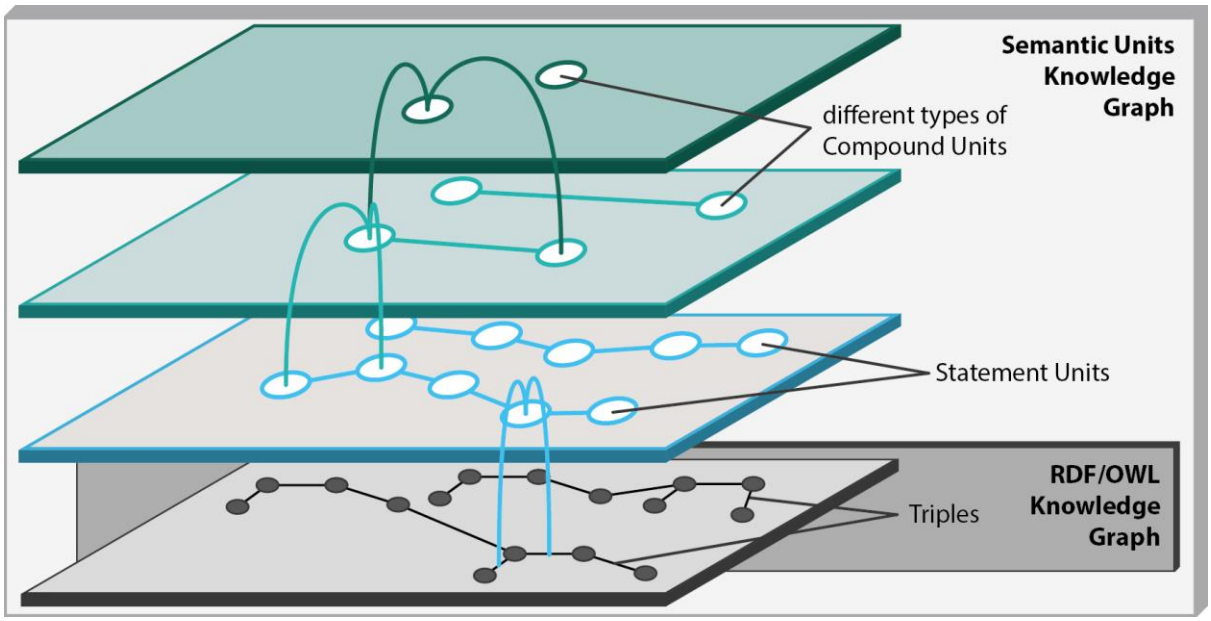


图1：语义单元在RDF/OWL三元组层之上，为知识图谱增添了更多层。三元组层

在数学上被划分为陈述单元层，使每个三元组恰好属于一个陈述单元，而每个陈述单元由一个或多个三元组组成。陈述单元可以组织成不同类型、语义上有意义的集合（即复合单元），借此可以定义各种附加层，以语义上有意义的方式进一步构建和组织知识图谱。

框 1 | 约定

在本文中，我们将 FAIR知识图谱称为机器可操作的语义图，用于记录、组织和表示断言性陈述（例如经验数据）

在本文中，我们用常规下划线表示本体类，用斜体下划线表示属性（即 Neo4j 中的关系），并用 ID 编号来

其中标签指类标签，ID 编号指类。此外，我们使用资源一词时，指的是具有唯一标识（例如统一资源标识符

为清晰起见，在正文和所有图表中，我们用人类可读的标签表示资源，而不是使用其 UPRI，并默认每个属

挑战 1：FAIR 经验数据必须明确说明建模所用的图模式，以避免模式互操作性冲突

人们通常认为，FAIR 意味着数据和元数据语句要实现可互操作和可重用，其中使用的所有概念都必须具有标识符，而这些标识符由本体等受控词汇表提供。人们常常忽略的是，要实现 FAIR，不仅概念必须标准化（即术语互操作性），数据和元数据语句中概念彼此关联的方式，以及这些语句所依据的**语义图模式**（即模式互操作性）也必须标准化——尤其是在知识图谱中存储和管理时。对于全称陈述，在基于RDF的知识图谱中，通常使用类公理中的 OWL 专用对象属性明确定义此类图模式（另见图 5B），从而形成 TBox 表达式。

对于经验数据，情况则有所不同。经验数据应建模为 ABox 表达式 (30)。然而，由于 RDF 和 OWL 具有很强的通用表达能力，在知识图谱中，任何给定的实证数据陈述都可以通过多种通常无法直接互操作的方式建模。机器很难识别出两个结构不同、但实际上描述同一底层数据陈述的 ABox 表达式。因此，只要数据是通过不同的图模式建模的，我们就必须处理此类模式互操作性冲突（参见图 2 和图 3）。

Observation:

ObjectX weighs 5 kilograms

Observation Graph:

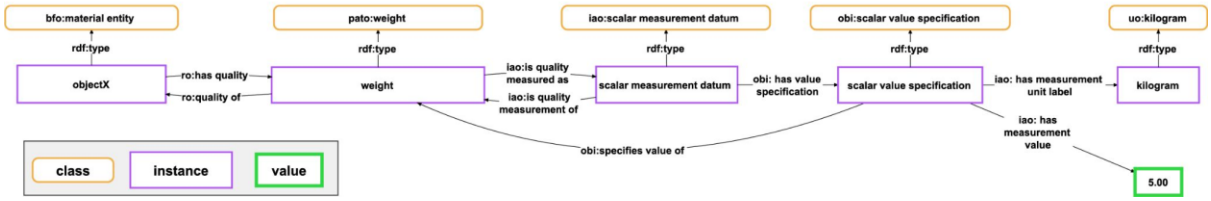


图 2：人类可读陈述与其遵循 RDF 语法、以 ABox 语义图表示的机器可操作形式的比较。 上图：关于观测的一个人类可读陈述，即 ObjectX 的重量为 5 千克。**下图**：同一陈述的图表示，使用 RDF，并遵循生物医学调查本体（OBI；http://obi-ontology.org/）（31）中测量数据的一般模式，该本体由开放生物学与生物医学本体库（OBO；http://www.obofoundry.org/）维护。

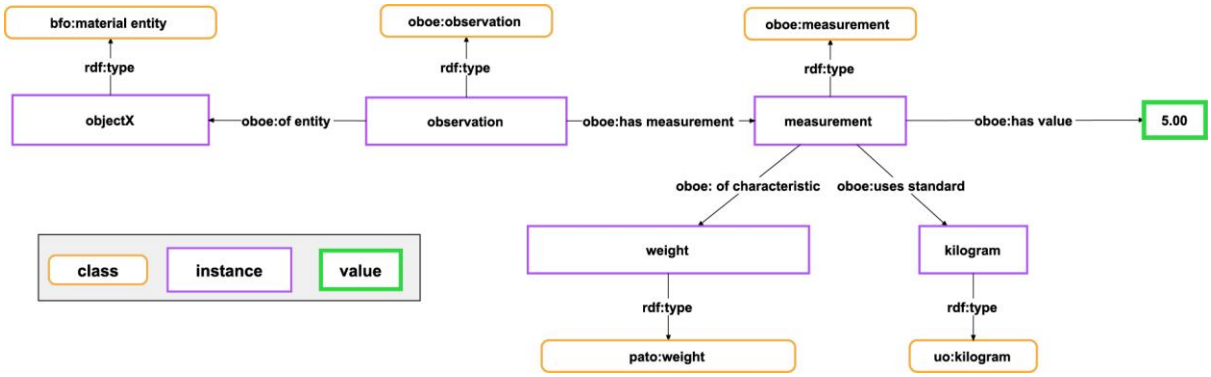


图 3：数据陈述的另一种机器可操作表示，源自图 2，遵循 RDF 语法和可扩展观测本体（OBOE）的图模型。 此图表示与图 2 上图所示相同的数据陈述，但采用了不同的语义图模型来表示，该模型基于 OBOE（http://bioportal.bioontology.org/ontologies/OBOE），这是生态学界经常使用的一种本体。

因此，要使实证数据陈述的 ABox表示符合 FAIR，就必须知道其语义建模采用了哪种图模式。只有采用相同模板、以图模式形式建模的基于实例的图（即三元组的主语 和宾语位置上为实例资源的图；ABox），才能保证满足互操作性的最低要求；例如，该模板可通过 SHACL 指定为一种形状 (32)。理想情况下，同一类型的语句（例如所有重量测量）都应使用相同的图模式，以便实现潜在的互操作性。因此，我们需要**此类图模式的标识符**；如果以 ABox 形式记录实证数据，其元数据就应指称相应的图模式标识符。借助这些信息，可以根据共同使用的图模式标识符，识别出可能互操作的 ABox 表达式。

实际上，这意味着：(i) 知识图谱中的所有语句都必须归入不同的语句类，每个类都应指定一个关联的图模式（例如，以形状规范的形式）；(ii) 必须能够识别属于特定语句的子图。只有满足这两个标准，数据和元数据的 ABox 表示才真正符合 FAIR指导原则。语义单元提供了一种满足这两个标准的方法。

挑战2：许多软件开发者看不到图查询语言的益处

大多数知识图谱要么是基于 RDF/OWL 并存储在元组存储中的有向标记图，要么是 Neo4j 之类的带标签属性图。要直接与这些图交互，也就是说，要对知识图谱中的语句执行创建（=写入）、读取（= 搜索）、更新和删除等 CRUD操作，就需要使用查询语言。对于 RDF/OWL，例如使用 SPARQL（https://www.w3.org/TR/rdf-sparql-query/）；对于 Neo4j，则使用 Cypher(https://neo4j.com/developer/cypher/)。

尽管这些查询语言支持详细且非常复杂的查询，但使用 SPARQL 或 Cypher 编写查询并非易事。知识图谱应用的用户通常缺乏自行编写此类查询所需的背景知识。遗憾的是，根据我们的亲身经验，即使大多数开发者也不熟悉这些语言，在尝试学习时也会为其复杂性所困扰。换言之，必须编写 SPARQL 或 Cypher 查询，构成了与知识图谱交互的入门门槛，也阻碍了其更广泛的应用（33）。因此，我们需要通过技术方案来缓解这一挑战，例如公开提供可重用的查询模式，并将其关联到特定的图模式。

语义单元的概念有助于实现此类解决方案：语义学专家可以为每种语义单元提供通用的 CRUD 查询，领域专家作为采用语义单元的知识图谱的用户和开发者，无需自行编写查询即可使用这些查询。不同类型语义单元的查询可以通过并集或交集进行组合，从而构成更复杂的查询。

挑战3：关于陈述的陈述可能颇具挑战性

主语、谓词和宾语 的 RDF 三元组语法本身并未提供一种良好的方法来表达关于陈述的陈述。在 RDF 中，关于某个陈述的陈述是一个三元组，它将由一个或多个三元组组成的陈述与某个数值、资源或其他此类陈述关联起来。

遗憾的是，如果你想在 FAIR知识图谱中表示经验数据，或例如学术出版物的内容，通常就必须对关于陈述的陈述进行建模。例如，若要为测量数据提供详细的元数据，实际上就必须将两个子图相互关联：一个是表示测量本身的图，由一组三元组记录（例如，见图 2）；另一个是记录测量过程的图，其中也包含若干三元组（例如，见图 4）。

如果我们只想在另一个三元组中指称单个三元组语句，可以使用 RDF 再述。在 RDF 再述中，会定义一个资源来表示特定三元组，并通过另外三个三元组描述它，分别指定其主语、谓词和宾语。或者，也可以使用 RDF-star 方法（34,35）。然而，在 RDF-star 中，所有涉及相同主语、谓词和宾语的语句都被视为指称同一个三元组，因此不会彼此区分；结果便是无法区分同一语句的不同实例。此外，虽然使用 RDF-star 指称单个三元组是可行的，但当需要指称多个三元组语句组成的集合，进而指称更大的子图时，查询会变得非常低效且复杂。如果要使用重ification对图4所示的整个图作出陈述，首先必须为图中的每个三元组分别指定一个语句。由于该图包含 20 多个三元组，光这第一步就需要创建 60 多个三元组。然而，由于你想指称的是整个图，而不是其中的每个三元组，因此还必须额外创建 20 多个三元组，将表示整个子图的资源与第一步中创建的、用于表示各个三元组的资源连接起来。虽然这在技术上可行，但既不简洁，也不便于查询。

在这种情况下，使用命名图比采用 RDF 再述或 RDF-star 高效得多。命名图可用于基于 RDF 的知识图谱。命名图资源通过将子图的 UPRI（从而也包括语句的 UPRI）作为第四个元素添加到每个三元组中，来标识一组三元组陈述，并将三元组转化为四元组。此外，在执行更复杂的查询时，命名图还有一项额外优势：其性能优于其他元数据表示模型 (36)。

另一方面，在带标签属性图中，为标识整体数据图内的子图而分配资源十分简单，只需将资源标识符作为相应属性—值对的数值，并将该属性—值对添加到属于同一子图的所有关系和节点即可。

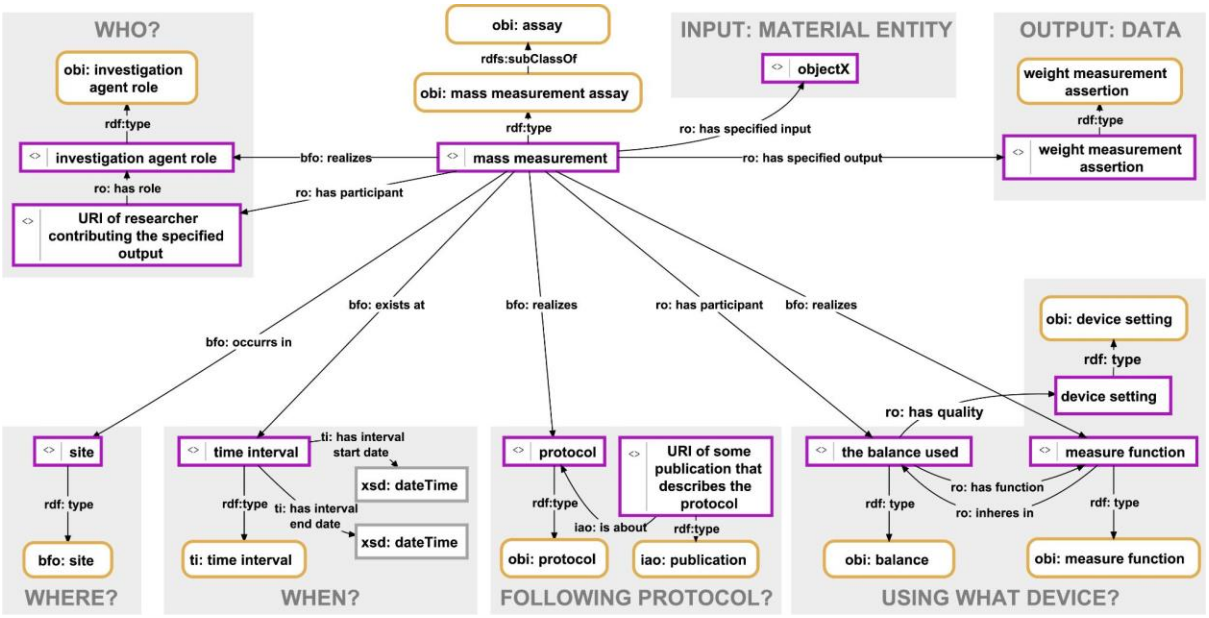


图4：一个 关于重量测量数据的元数据的详细机器可操作表示

据记载为 RDF ABox 图。该表示采用遵循 RDF 语法的 ABox 语义图形式。该图记录了使用天平进行的质量测量过程。它将一个质量测定实验（OBI:0000445）的实例与来自不同本体的其他各类实例关联起来，说明了测量由谁执行、在何处及何时进行、遵循了哪项规程以及使用了哪种设备（即天平）。此外，该图~~还指定了作为主语、因~~而作为测量过程输入的特定物质实体（即“*objectX*”），并指定了作为观测输出、包含在特定重量测量断言中的数据。

将整体数据图组织成不同的语义单元，从而形成处于不同表征粒度层级、具有语义意义的子图，是一种高效的图结构化方式，使用户能够直观地对陈述作出陈述；同时，这也提供了清晰、直接的实现模式，有助于开发相关的搜索模式。

挑战4：RDF/OWL 与属性图之间的概念鸿沟，以及区分全称、偶然性和断言式陈述的问题

有些知识图谱基于 RDF/OWL，另一些则基于带标签属性图。每种技术都有其特定的优势和不足（见表1）。

表1：基于 RDF/OWL 和带标签属性图的知识图谱的优缺点比较 (37,38).

	RDF	**带标签属性图**
查询语言	SPARQL，万维网联盟建议标准	多种，没有一种是万维网联盟 建议标准
形式语义和 推理	OWL；然而： 1. 映射到 RDF 时，OWL 通常会非常 冗长，导致不必要的复杂性 无法直观理解的图 人类读者难以理解，并且 会妨碍有效计算； 2. OWL 并非唯一的推理模式， 除描述逻辑之外，还存在其他 用于推理的框架	尚无统一的标准体系（没有万维网联盟 建议标准）——标准化的 OWL 到属性图的映射 仍然缺失；Neo4j 的数据模型 并非基于形式化的集合论或 一阶逻辑
关联信息 直接关联到边	只能通过重ification、RDF-star 或 命名图（见 挑战*3*）	信息可以直接关联到 关系，即边
全称的区分 ， 偶然和断言式 陈述	……之间的形式化语义区分 断言式陈述与全称陈述； OWL 中未对偶然陈述建模	没有形式化的语义区分

基于 RDF 的图与带标签属性图之间的一个主要区别是，OWL 可以映射到 RDF，因此能够依据万维网联盟的一项建议标准提供形式语义和推理。目前，尚无针对 Neo4j 等属性图的万维网联盟推荐标准映射，尽管已有一些提案（例如 OWLStar、owl2lpg）。OWL 基于描述逻辑，因此严格区分**全称陈述**[1]（以**类规范**的形式表示；术语盒，**TBox**）与**论域**，后者以**断言性陈述**[2] 的形式表示，这些陈述将实例（即个体）彼此关联起来（断言盒，**ABox**）（38）。由于 OWL 中的类具有 URI，因而可以被引用，而它们的类公理却没有 URI，因此 TBox 不属于论域——在知识图谱中，无法引用某个类的特定公理，只能引用类本身。因此，无法针对类

[1] 全称陈述是指对某个特定全称的每一个实例都成立的语句。全称陈述代表普遍接受的领域知识，并说明必然成立的内容。全称陈述基于多对一关系。本体类的定义及其类公理是全称陈述的示例，许多科学假设和类定律规律也采用全称陈述的形式。[2] 断言式陈述是指对特定个别事物成立并指出某个事实的语句。断言性陈述基于一对一关系。在知识图谱中，断言性陈述将特定实例彼此关联起来。经验数据是断言性陈述的示例。

RDF/OWL 知识图谱中的定义（即全称陈述）。因此，如果全称规律*X*被建模为一个类，那么在知识图谱中记录诸如“作者 *A* 断言全称规律 *X*”这样的陈述并非易事。

另一方面，由于属性图通常不区分 ABox 和 TBox 表达式，因此其语句的实际含义并不总是明确。例如，语句“手具有部分拇指”（图 5A）可能指某只特定的手及其拇指，例如本文某位作者的右手。如此一来，该语句便代表一个断言式陈述。不过，这句话也可能指某只未知的手，意为至少存在一只手，其部分是某个拇指；这代表一个**偶然陈述**[3]。对此语句的另一种可能解读是，它记录了与手相关的一种典型情形：手通常具有拇指。在这种情况下，该语句代表一个**原型陈述**[4]，，它是偶然陈述的一种特殊情况。最后，该语句也可能指类“手”（FMA:9712）的所有实例，并代表一个全称陈述，表明每只手必然都具有拇指。换言之，由于缺乏形式语义，带标签属性图没有标准化的方法来区分全称陈述、偶然陈述、原型陈述和断言性陈述。

关于全称陈述，从逻辑角度来看，通过属性 具有部分(BFO:0000051) 将手（FMA:9712）和拇指（FMA:24938）这两个类联系起来毫无意义，因为类不能以其他类为其部分。除少数例外（例如，*subclassOf*（RDFS:subclassOf）、类型（RDF:type））外，对象属性所建模的是实例与实例之间的关系，而非类与类或实例与类之间的关系。尽管如此，如果仍然使用这类对象属性来描述类之间的关系，就会导致语义不一致。这或许不会给人类读者造成问题，但会给推理机等机器带来问题。

在 OWL 中，全称陈述‘每一个实例手都有某实例属于拇指作为其部分’可以通过 OWL 特定属性以语义恰当的方式表达，从而得到一种复杂但语义精确的陈述表示（图 5B）。然而，人类读者通常不想查看这种复杂的表示，而更偏好属性图所提供的简单且仍然直观易懂的表示（图 5A）。此外，基于 OWL 的

[3] 偶然陈述是指对于特定全称的某些实例为真，并说明可能存在的情况的陈述。偶然陈述基于部分到部分关系。在知识图谱中，偶然陈述对于某个类的部分实例为真，但不一定对每一个实例都为真。[4] 原型陈述是指对于特定全称的所有实例都为真（除非明确说明情况并非如此），因此说明的是通常成立但不一定成立的情况。

表示涉及空白节点，这使得基本的 SPARQL 查询难以执行（38），并可能引发其他问题（39）。

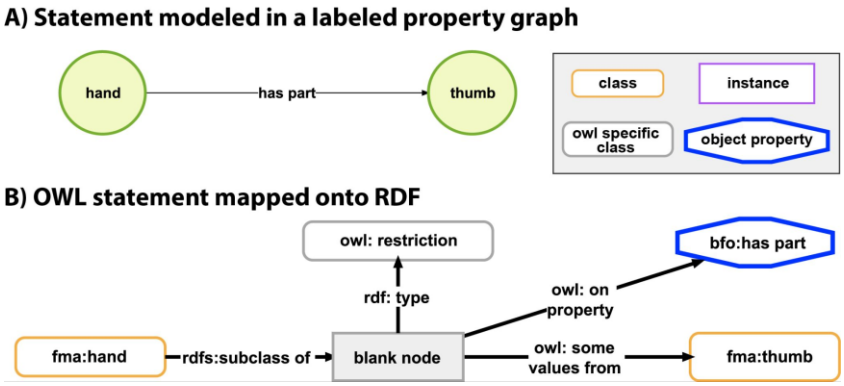


图5：属性图与映射到 RDF 的 OWL 对全称陈述的表示方式对比。该陈述'每一个手的实例都具有某个拇指作为部分'在以下形式中的表示：**A)**带标签属性图；**B)**映射到 RDF 的 OWL。

已有研究提出了在属性图中表示全称陈述的方案。这些方案通过为边添加逻辑属性（例如存在限制公理）来实现，从而形成对象属性映射（37,38）。然而，目前仍缺少万维网联盟关于将 OWL 映射到属性图的形式化方法的标准化建议标准。

将知识图谱组织成不同的语义单元，每个单元都有自己的 UPRI，并被归类为全称陈述、偶然陈述或断言式陈述，不仅能提供正式的概念框架，便于应用这类对象属性映射，还能针对全称陈述作出陈述，从而使全称陈述可供论域使用。这意味着正式的类定义（即类公理）将变得可指称，因此可以在图中对其作出陈述。

结果

将知识图谱组织成语义上有意义的表示单位

为了使知识图谱更易于管理，并赋予其更高程度的结构化特征，必须将图组织为多个部分重叠、部分封闭的子图层。此前已有将知识图谱系统地划分为子图的方案，例如，将**特征集**定义为由所有在主语 位置共享同一资源的三元组构成的子图，从而显著提升空间利用率和查询性能（40,41）（另见相关概念**RDF 分子**（42,43））。

然而，我们不采用基于三元组共享的图拓扑属性这一较为技术性的方面来将知识图谱划分为子图，而是提出一种方法，重点在于**将知识图谱组织为可识别且具有语义意义的三元组集合，即表示单元的子图（简称：语义单元）**。该方法基于语义数据模型领域提出的一种构想：将表型描述组织为最小的语义单元，即断言，并将每个断言置于各自的命名图中（44）。在此，我们将这一构想推广并扩展为语义单元的概念。这些语义单元可作为可识别、可重用的数据项独立访问、搜索和重用，构成表示单元；它们是 FAIR数字对象，可使用基于RDF/OWL的图以及属性图实现。

语义单元是整体知识图谱的一个子图，表示**对人类读者而言具有语义意义的信息单元**。每个语义单元在图中都由其自身的资源（即节点）表示，因此拥有自己的 UPRI。此外，它还可以归类为相应本体类的实例。

在由语义单元构成的知识图谱中，可以区分两个不同的图层：数据图层和语义单元图层。**数据图层**包含知识图谱中的所有三元组，语义单元将这些三元组组织成可识别的子图。未由语义单元构成的知识图谱仅包含数据图层。将知识图谱组织为语义单元时，会为其添加**语义单元图层**，其中包含因将图谱组织并构建为语义单元而添加到图中的所有三元组。每个语义单元的资源及其相互关系，例如，都属于语义单元图层。与知识图谱整体的这种区分类似，我们也可以为每个语义单元区分数据图和语义单元图。特定语义单元的数据图与其语义单元资源拥有相同的 UPRI，因此，引用该 UPRI 时，既指向作为资源的语义单元，也同时指向其对应的数据图，从而可以对数据图的内容作出陈述（图 7）。

语义单元可以分为两个基本类别及其子类别：**陈述单元**和**复合单元**（语义单元的分类见图 6 ）。

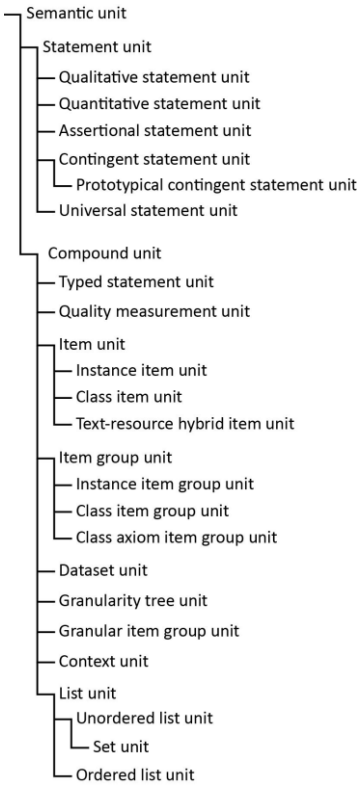


图6：不同类别语义单元的分类。

陈述单元

陈述单元的定义是：**对人类读者而言，在语义上有意义、能够表示最小独立命题（即陈述）的信息单元**（另见（44））。例如，某个人群感染了病毒，并具有基本再生数这一质量（NCIT:C173777），这一事实表示一个陈述单元；它独立于基本再生数的实际测量，而后者表示另一个陈述单元。测量陈述单元独立于质量陈述单元，因为由于测量误差、测量不准确性，或应用了不同的测量方法、仪器或模型，同一人群在同一时间点的基本再生数可能有多次测量。所有质量测量也都如此。例如，图 2 中*ObjectX*的重量测量由两个陈述单元组成。一个陈述单元说明*ObjectX* 具有"重量"这一质量（PATO:0000128），另一个则说明该重量的测量值。

将知识图谱组织成陈述单元，会形成其**数据图层的数学划分**，其中数据图层的每个三元组都恰好属于一个陈述单元。需要注意的是，只有数据图层会被划分为更小的数据（子）图。所有

为将知识图谱组织成语义单元而添加到知识图谱中的三元组，即语义单元图层中的三元组，都被排除在该划分之外；否则，应用语义单元将导致无限回归，因为添加一个语义单元总会向语义单元图层添加三元组，而如果将这些三元组纳入划分，就必须通过一个陈述单元来表示它们自身，这又会再次向语义单元图层添加三元组，如此循环往复。

遵循语言学中的**谓词-论元结构**，陈述中的主要动词及其助动词共同表示该陈述的谓词。每个谓词都有一个**配价**，它决定了谓词需要多少个以及何种类型的**论元**才能表达完整的意义。**附加语**也可以与谓词相关联，但它们并非表达谓词完整意义所必需的，只提供可选信息，例如具有部分关系陈述中的时间说明。主语和宾语短语是最常见的论元和附加语。因此，我们可以说，每个陈述单元都通过将某个资源与某个字面量或另一资源相关联，来记录一个特定的命题；该资源是命题谓词的主语论元，而字面量或另一资源则是该谓词的宾语论元或宾语附加语。我们将陈述单元中命题的主语论元称为**主语**，将宾语论元和宾语附加语称为陈述单元的**宾语**。每个陈述单元都有一个这样的主语和一个或多个宾语。

例如，一个具有部分关系陈述单元（图 7）有一个主语和一个作为论元的宾语，而一个重量测量陈述单元有一个主语和两个作为论元的宾语，即重量值和重量单位。图中表示陈述单元的资源通过属性 **hasSemanticUnitSubject** 与其主语相关联；这一属性记录在语义单元图中。

如果数据图中记录的命题基于**二元关系**，因而具有二价谓词，例如“拉尔斯的右手具有部分拉尔斯的右拇指”，那么相应的陈述单元通常由一个三元组构成，因为在 RDF 中，三元组的属性是二元关系。然而，许多命题基于**多元关系**，因此无法用单个三元组建模。例如，“数值为 *5*、单位为千克的重量测量”和“拉尔斯的右手于一月具有部分拉尔斯的右拇指 29^{th} *2022*”这两条语句至少由两个三元组构成，而“*Anna* 给 *Bob* 一本书”“*Carla* 于 *2022* 年 *6* 月 *29* 日从巴黎乘火车前往柏林”和“*Dave* 从 *Eric* 那里买下这辆车”则各自需要两个以上的三元组。因此，陈述单元可以由一个或多个三元组构成，具体取决于底层命题所涉及的关系及其谓词的配价，也取决于需要额外建模的可选附加语数量。因此，可以参照对底层关系进行建模的多元属性，定义一种特定的陈述单元类型，即陈述单元类。许多对象

基础形式本体 2 的许多属性实际上是三元关系，因为它们具有时间依赖性（45）。例如，在 RDF 中对“*b* 位于_*c*，时间为 *t*”进行建模，至少需要两个三元组。

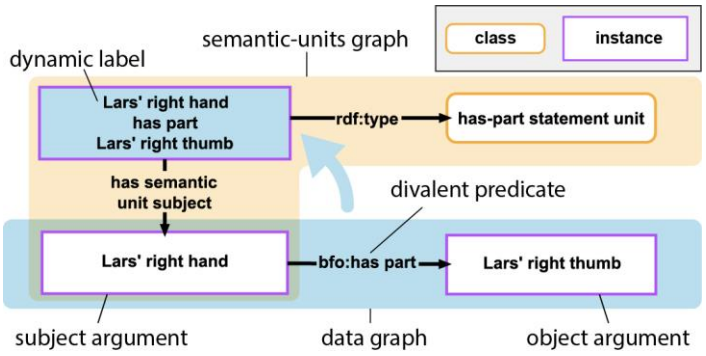


图 7：陈述单元示例。 一个语义单元示例，它基于两个实例之间的部分关系来建模一个陈述。表达该陈述的数据图显示在底部的蓝色框中，其中“Lars 的右手”是该陈述的主语参数，“Lars 的右拇指”是宾语参数。语义单元图显示在桃色框中，其中包含表示该语义单元的三元组。它指定表示陈述单元的资源（带边框的蓝色框，此处显示其动态标签）是部分关系语句单元的一个实例，并且该语义单元的主语是“Lars 的右手”；该陈述单元表示数据图中的陈述（由蓝色箭头指示）。部分关系语句单元的 UPRI 也是该语义单元数据图（蓝色框中无边框的子图）的 UPRI。

换言之，哪些三元组属于一个陈述单元，必须由人类领域专家针对每种底层谓词类型逐一指定。然而，通过指定图模式，将给定命题建模为基于图的表示，而该表示又构成一个陈述单元，可以使用诸如LinkML或形状约束语言SHACL (32) 之类的建模语言，并可应用相应工具来识别此模式的实例，从而识别相应陈述单元类型的实例。此外，在用户界面 (UI) 中将这些单元显示为知识图谱的可视化时，可以利用 SHACL 形状为每个陈述单元生成**动态标签**。此类标签可通过解析陈述单元主语资源的标签以及其所有宾语资源和字面量的标签，并根据预定义的模板将它们组合成人类可读陈述，从而动态生成。例如，模板“主语乘坐*A*从*B*到*C*于*D*”，输入“*Carla*| 火车| 巴黎| 柏林| *29*月”后，生成的陈述为“*Carla* 乘坐火车从巴黎到柏林于*2022*年*6*月*29*日”。

每个陈述单元都实例化相应的陈述单元本体类。该类的特征是其实例所建模的命题的关系类型。根据其底层命题的关系，可以区分陈述单元的两个子类别（**基于关系的标准**）：

1. **定性陈述单元**：记录定性关系的陈述单元，例如具有部分关系（图 7）。**定性陈述单元的论元对象始终是**

资源而非数值——只有定性陈述单元的附加宾语可以是数值。

2. **定量陈述单元**：用于记录定量陈述的陈述单元，例如测量数据。**定量陈述单元的至少一个论元对象必须是数值**。

这两个类别既穷尽所有情况，又互不重叠，这意味着任何给定的陈述单元要么是定性陈述单元，要么是定量陈述单元，但绝不会两者兼具。在这两个子类别中，可以根据其底层关系（例如具有部分、类型、发展自）进一步区分不同的陈述单元类。给定关系类陈述单元的其他子类型，例如关于物质实体的具有部分关系陈述与关于信息内容实体的具有部分关系陈述，无需作为不同的类加以区分，而可以通过查询具有部分关系陈述单元类的实例与其主语之间的 **hasSemanticUnitSubject** 关系来识别和区分。

除了基于关系的判据之外，还有另一个区分陈述单元类的判据。**基于主体类别的判据**依据陈述单元主语所属的实体类别，将陈述单元区分为断言式、偶然式和全称陈述单元。这三个陈述单元类别既穷尽所有情况，又互不重叠，这意味着任何给定的陈述单元要么是断言式陈述单元，要么是偶然陈述单元，要么是全称陈述单元，但绝不会是它们的组合（另见挑战*4*）。由于基于关系的判据与基于主体类别的判据彼此独立，给定的陈述单元例如可以同时是重量测量陈述单元和断言陈述单元。

在讨论基于主语类别的不同类型陈述单元之前，我们想先介绍一种约定，并将其应用于语义单元，以帮助弥合挑战*4*中讨论的概念鸿沟。此约定涉及在实例、属性和类之外引入两种新类型的资源：**某个实例资源和每个实例资源**。基于这两种新类型的资源，可以区分出三种不同类型的定性陈述单元，它们对数据图中某个实例、每个实例、命名个体和类资源之间的三种不同类型关系进行建模。

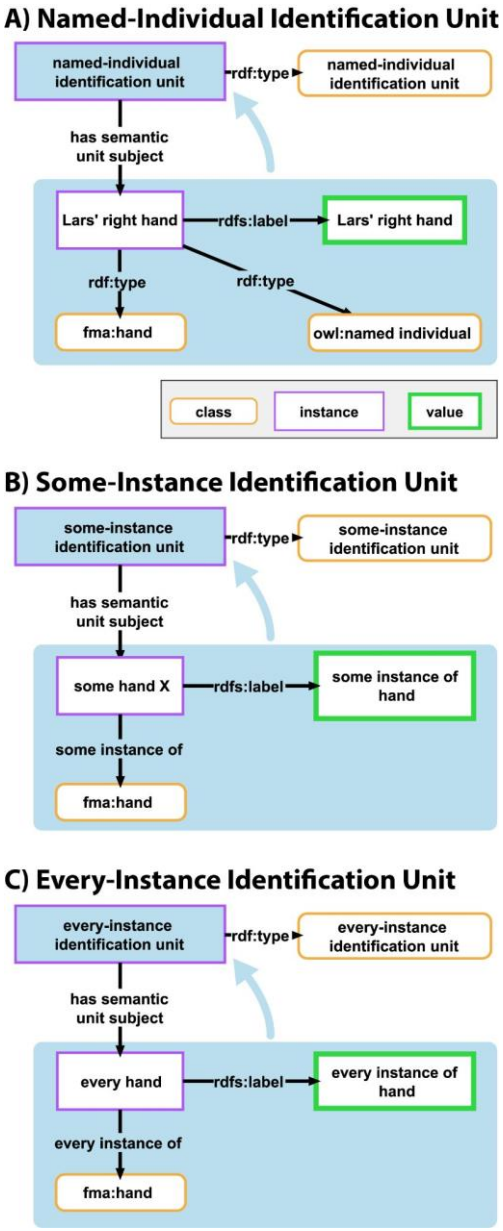


图 8：三种不同类型识别单元的示例。A）命名个体识别单元。该命名个体识别单元在其数据图中，通过属性类型（RDF:type）对作为主语的实例"拉尔斯的右手"（FMA:9712）的类别归属进行建模，并通过属性标签（RDFS:label）记录主语的标签"拉尔斯的右手"。无边框蓝色框表示属于该命名个体识别单元（带边框蓝色框）的数据图。命名个体识别单元的主语是实例的类型，这些实例也作为断言陈述单元的主语（参见图 9A）。**B）某实例识别单元。**该某实例识别单元在其数据图中，通过属性*某实例属于*对作为主语的实例"某只手 X"（FMA:9712）的类别归属进行建模，并记录主语的标签"某个实例属于手"。无边框蓝色框表示属于该某实例识别单元的数据图。某实例识别单元的主语是实例的类型，这些实例也作为偶然陈述单元的主语（参见图 9B）。**C）每个实例识别单元。**该每个实例识别单元在其数据图中，通过属性*每一个实例属于*对作为主语的实例"每只手"（FMA:9712）的类别归属进行建模，并通过属性标签（RDFS:label）记录主语的标签"每个实例属于手"。无边框蓝色框表示属于该每个实例识别单元的数据图。每个实例识别单元的主语是实例的类型，这些实例也作为全称陈述单元的主语（参见图9C）。

1. **命名个体识别单元**：一种定性陈述单元，用于识别陈述单元中作为主语的资源，将其指认为命名个体，并包含其标签和类型，即其类成员资格（图 8A）。

2. **某实例识别单元**：一种定性陈述单元，它标识一个资源，该资源是陈述单元的主语，指代给定类的某些未指定实例，并附有标签及其类规格（图8B）。当此资源用于某个其他陈述单元时，它表示'∃i∈C'（其中 i = 是实例，C = 是类），读作'存在类 C 的一个实例 i，使得……成立'。此外，某个实例识别单元还可以包含有关基数限制的信息（见下文的否定建模、基数限制和分歧）。

3. **每个实例识别单元**：一种定性陈述单元，它标识一个资源，该资源是陈述单元的主语，指代给定类的每一个实例，并附有标签及其类规格（图8C）。当此资源用于某个其他陈述单元时，它表示'∀i∈C'（其中 i = 是实例，C = 是类），读作'对于类 C 的每一个实例 i，..... 成立'。

断言陈述单元

断言式陈述单元是指以**命名个体**作为其主语资源的陈述单元（例如图 8A 中的"拉尔斯的右手"（FMA:9712）），因此其包含一个知识图谱已知的实例。换言之，它的主语也是一个命名个体识别单元的主语。如果断言式陈述单元的宾语是资源而非字面量，那么该宾语也指代一个命名个体，从而形成一对一关系（对于 n 元命题，则会形成多个此类关系）。因此，断言式陈述单元是至少涉及一个特定个体的关系的 ABox 表示。

作为一种语义单元，特定的断言式陈述单元指向知识图谱中对特定断言进行建模的子图。断言是关于特定个体（主语）的命题，因此是关于个体而非全称的命题；该命题旨在表达真或假（46,47）。在知识图谱中，特定的断言式陈述单元由一个表示该单元数据图的资源建模。因此，该数据图是知识图谱中基于实例的子图，并表示一个**ABox表达式**。单元资源既是相应陈述单元类型的实例，也是*断言式陈述单元*的实例（图 9A）。

或然陈述单元

或然陈述单元是一种以**特定类的某些实例**为主语的陈述单元。换言之，它的主语也是某实例识别单元的主语。如果或然陈述单元具有一个作为资源的宾语，那么该宾语也指向特定类的某些实例，从而形成某对某关系（若为 n 元命题，则会形成多个此类关系）。例如，如果你想表达某些手（但不一定是全部手）有拇指，而又不想或无法列出所有适用的个体手，就可以使用这种或然陈述单元（图 9B）。

偶然陈述是指对于某个全称的部分实例为真、但未必对每个实例都为真的命题（46–48）。它们常用于表达**原型关系**，例如人类的手通常有一个拇指作为其部分，因此，这类陈述谈论的是在特定语境中被认为通常为真的全称，但并非必然如此——

即使手没有拇指，它仍然是手[5]。其他例子包括关于疾病与其症状之间的关系，或药物与其作用之间的关系的陈述；这些关系通常通过某种概率评估来表述。这些原型偶然陈述可以记录在**原型或然陈述单元**中，后者构成偶然陈述单元的一个子类别。

因此，描述逻辑以及 OWL 都无法表达通常为**真**的内容；要为这类语句（47）提供语义，还需要能够表达原型知识的其他逻辑形式体系。

在知识图谱中，特定的或然陈述单元可建模为一个陈述单元，其资源表示相应的数据图。该资源既是相应陈述单元类型的实例，也是*或然陈述单元*的实例。其主语以及（如有）宾语均建模为资源，并通过相应某实例识别单元中的 *某实例属于* 属性与其所属类建立关系（图8B）。表达关于某陈述为真 时实例频率的假定数值的概率值，可通过附加三元组记录；这些三元组进而构成独立的断言陈述单元，其主语为或然陈述单元的资源。

[5] 因此，图9C中的示例并非人体解剖学家会作出的陈述。他们会通过其他属性来指称人手。

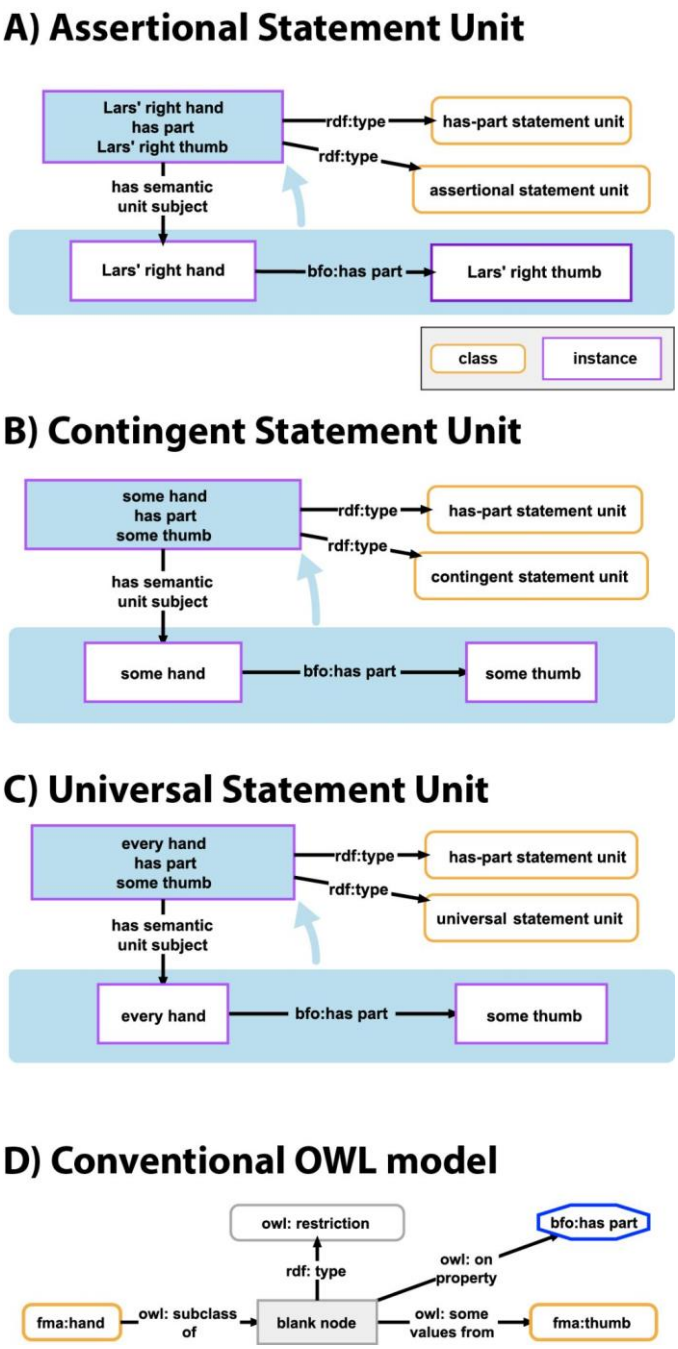


图9：陈述单元的三个子类别。 A）断言陈述单元。断言陈述单元的示例。它对两个实例之间的具有部分关系进行建模。表达该关系的数据图显示在无边框的蓝色框中。它是基于实例的数据图，构成一个 ABox。此断言陈述单元的主语是一个实例资源，同时也是一个命名个体识别单元的主语（参见图8A）。表示该数据图的资源，即断言陈述单元资源（带边框的蓝色框，此处显示其动态标签），是*具有部分关系陈述单元*和*断言陈述单元*的实例；这一点由无边框蓝色框上方的语义单元图表示。**B）或然陈述单元。**或然陈述单元的示例。它对一种具有部分关系进行建模，该关系存在于“手”（FMA:9712）的某些实例（但不一定是每一个实例）与“拇指”（FMA:24938）的某些实例（但不一定是每一个实例）之间。表达该关系的数据图显示在无边框的蓝色框中。此或然陈述单元的主语是一个实例资源，同时也是一个某实例识别单元的主语（参见图8B）。表示该数据图的资源，即或然陈述单元资源（带边框的蓝色框，此处显示其动态标签），是*具有部分关系陈述单元*和*或然陈述单元*的实例。**C）全称陈述单元。**一个一个陈述单元的示例，它是一个*全称陈述单元*。它描述了“手”（FMA:9712）的每一个实例与“拇指”（FMA:24938）的某些实例（但不一定是每一个实例）之间存在的具有部分关系。表达这一关系的数据图显示在无边框的蓝色框中，原则上可以将其转换为 OWL，以形成 TBox。这个全称陈述单元的主语是一个实例资源，同时也是一个每个实例识别单元的主语（参见图8C）。表示这一数据图的资源，即全称陈述单元资源（带边框的蓝色框，此处显示其动态标签），是*具有部分关系陈述单元*和*全称陈述单元*的实例。**D) 传统的 OWL 模型。**将 C) 中的示例转换为映射到 RDF 的 OWL 类公理表达式。

全称陈述单元

全称陈述单元是一种陈述单元，其主语指向**某个特定类的每一个实例**。换言之，它的主语也是一个每个实例识别单元的主语。如果全称陈述单元具有一个作为资源的宾语，该宾语便指向某个特定类的若干实例，从而形成一种全到部分关系（若为 n 元命题，则会形成多个此类关系）。如果要陈述“手”（FMA:9712）的每一个实例必然具有部分（BFO:0000051）某个“拇指”（FMA:24938），就要使用全称陈述（图9C）。全称陈述表示普遍接受的领域知识，例如以术语定义（表达为类公理）的形式，或以关于规律性因果关系的假设的形式呈现。

作为一种语义单元，特定的全称陈述单元指的是知识图谱中的一个数据图，它表示一个特定的全称陈述，因而表示一个对特定全称的所有实例都为真的命题 (46,47)。在知识图谱中，特定的全称陈述单元被建模为一个陈述单元，其资源表示相应的数据图。该资源既是相应陈述单元类型的实例，也是*全称陈述单元*的实例（图9C）。其主语被建模为一个资源，该资源在相应的每个实例识别单元（图8C）中通过*每一个实例*属性与其类相关联；而其宾语资源（若存在）则被建模为资源，在相应的某实例识别单元（图8B）中通过*某实例属于*属性与其类相关联。

与植根于描述逻辑的全称陈述 OWL 建模方式（参见图9D）相比，作为语义单元的全称陈述单元所对应的数据图更简单，也不包含任何空白节点。这改善了其查询性能。此外，陈述单元的资源可用于对全称陈述作出陈述，从而使全称陈述以及特定类公理也能纳入一般论域（参见挑战*4*）。然而，由于目前尚无推理机能够直接基于这种语义单元表示法进行推理，因此若要对这些陈述进行推理，就必须将它们转换为基于 OWL 的表示法，形成经典的**TBox 表达式**（参见语义单元的逻辑语义）。

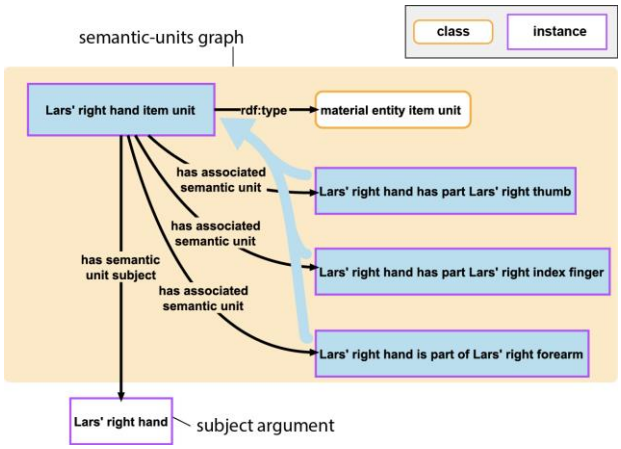
复合单元

复合单元是语义单元，它们将给定知识图谱中的所有陈述单元组织成更大但仍具有语义意义的数据图。从技术上讲，复合单元是特定语义单元资源的容器，因此定义了一个关联语义单元集合。合并其关联语义单元的数据图，即构成

复合单元的数据图。与陈述单元类似，每个复合单元都是一个语义单元，在图中由其自身的节点表示，拥有自己的 UPRI，并实例化相应的复合单元本体类。

在复合单元的语义单元图中，表示复合单元的资源与表示其关联语义单元的资源之间的关系，通过属性 **hasAssociatedSemanticUnit** 进行记录（图10）。复合单元可分为以下子类别。

图10：由多个陈述单元构成的复合单元示例。复合单元仅通过合并其关联陈述单元的数据图，间接拥有数据图。不过，复合单元资源（此处为 *'Lars的右手物品单元'*）代表这些合并后的数据图（蓝色箭头所示）。复合单元拥有语义单元图（显示在桃色方框中），用于记录与其关联的语义单元。



类型化陈述单元

为了让人类能够读懂陈述单元，必须将其主语、谓词和宾语的 URI 转换为相应的标签。陈述单元的动态标签正是以结构化的方式完成这一转换，将陈述本身呈现为人类可读的通俗英语句子。不过，动态标签所包含的信息超出了陈述本身，因为它还纳入了与陈述的主语及其宾语资源关联的识别单元中的信息。由于识别单元可以独立于将其资源用作主语和宾语的陈述单元进行编辑，因此，提及一条人类可读陈述时，实际上指的是相应的陈述单元，以及其主语和所有宾语各自的识别单元。这一语义单元集合就是类型化陈述单元。类型化陈述单元是由以下陈述单元构成的复合单元（图11）：

1. 一个不属于三种识别单元（见图 8）中任何一种的陈述单元。该陈述单元随后充当类型化陈述单元的**参照陈述单元**，其主语也就是与之通过属性 **hasSemanticUnitSubject** 相关联的类型化陈述单元的主语。

2. 对参照陈述单元的数据图中所有被指称资源的类别归属进行建模的**识别单元**。

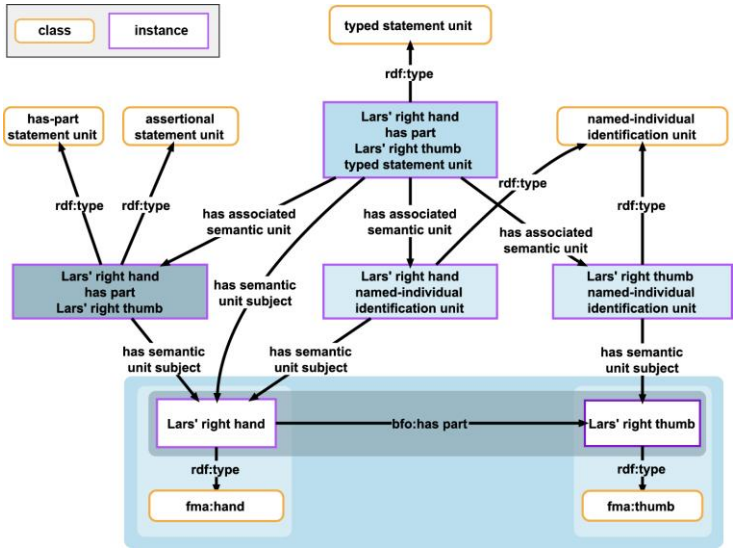


图 11：类型化陈述单元。类型化陈述单元的一个示例，数据图显示在无边框的大型蓝色方框中。该图由三个陈述单元的子图合并而成：来自图 9A 的*“部分关系断言语句单元”*，它是该*“类型化部分陈述单元”*的参照陈述单元，其数据图显示在无边框的灰色方框中；以及两个*命名个体识别单元*实例，它们的数据图显示在无边框的浅蓝色方框中（表示不同数据图的资源显示在相应颜色的带边框方框中，并带有动态标签）。

边框和动态标签）。

类型化陈述单元非常重要，因为其中包含类别归属和标签信息，而人类读者通常需要这些信息来理解参照陈述单元所表达的命题。由于表示类型化陈述单元的资源通过属性 *hasAssociatedSemanticUnit* 与其参照陈述单元及所有识别单元相关联，因此只需查询这些关系，就能识别不同类型的类型化陈述单元，无需进一步将类型化陈述单元这一类细分为子类。

质量测量单元

质量测量单元包含一个定性类型化陈述单元，以及一个或多个通过属性 *hasAssociatedSemanticUnit* 关联的定量类型化陈述单元。其定性类型化陈述单元的参照陈述单元用于指定给定实体的一项质量，而其每个定量类型化陈述单元的陈述单元则用于指定对该质量的测量，包括一个数值和一个单位。因此，质量测量单元是由一组类型化陈述单元构成的集合，这些陈述单元描述给定实体的一项特定质量，以及当前数据图中与之相关的所有定量评估或测量（图 2 所示的图即为质量测量单元典型的数据图）。

例如，某一人群在特定时间的基本再生数可以记录在这样的质量测量单元中，其中包括一个类型化定性陈述单元，用于关联该人群与质量基本再生数（NCIT:C173777）

以及一个或多个类型化定量陈述单元，每个单元都记录对该再生数的一次测量，其估算可能基于不同的方法和算法。

基本再生数测量陈述单元以质量陈述单元的宾语资源为宾语，并通过 *由语义单元描述的对象* 属性将质量陈述单元与基本再生数测量陈述单元关联起来。质量陈述单元则将其关联的质量测量单元的主语作为自己的主语，并通过 **hasAssociatedSemanticUnit** 属性与相应的质量测量单元关联，基本再生数测量陈述单元也是如此。

质量陈述单元的主语也是质量测量单元的主语，而语义单元图中表示质量测量单元的资源通过 **hasSemanticUnitSubject** 属性与该资源相关联。

由于语义单元图中表示质量测量单元的资源通过 **hasAssociatedSemanticUnit** 属性与其关联的类型化陈述单元及其参照陈述单元相关联，因此只需查询这些关系即可识别不同类型的质量测量单元，无需再将质量测量单元类细分为多个子类。

项目单元

项目单元由所有具有相同主语的陈述单元、类型化陈述单元和质量测量单元组成。相应的资源也因此成为**项目单元的主语**，而在语义单元图中表示项目单元的资源通过属性 **hasSemanticUnitSubject** 与其主语关联。因此，项目单元是语义相关的陈述单元、类型化陈述单元和质量测量单元的集合，它们都包含有关同一实体的信息。从概念上说，项目单元与按资源划分图~~数据管理模式 (49) 或~~前述特征集或*RDF* 分子相关，但它将这一理念应用于陈述单元而非三元组，类似于 Wikibase 数据模型中的项目。

项目单元中包含的信息很适合在用户界面的同一页面上呈现，因为其中包含的陈述单元、类型化陈述单元和质量测量单元描述的是同一实体，即它们共有的主语。项目单元可区分为以下子类别：

1. **实例项目单元**：主题为**命名个体**资源的项目单元。因此，它仅由知识图谱语义单元图层中所有具有相同主语的断言陈述单元组成。其主语也

必然是命名个体识别单元的主语。例如，关于某个特定受感染人群的所有语句，如该人群的位置、质量、各种测量数据（包括基本再生数和病死率），以及这些人群信息所指的时间段，都是共享同一主语的类型化陈述单元和质量测量单元，而该主语是一个命名个体。因此，它们共同构成一个实例项目单元。实例项目单元所含的信息可以在同一个用户界面页面中呈现。另一个例子是对某个特定标本解剖结构的描述：可以将描述该标本某个特定部分的所有类型化陈述单元和质量测量单元归入一个实例项目单元，由此形成多个这样的实例项目单元——每个被描述的部分对应一个。

2. **类项单元**：主体为某个实例资源或**特定类的每个实例资源**的项目单元。因此，它只包含知识图谱语义单元图层中主体相同的所有全称陈述单元和所有偶然陈述单元。图12所示的简单星形类公理可以组织成这样的类项单元。

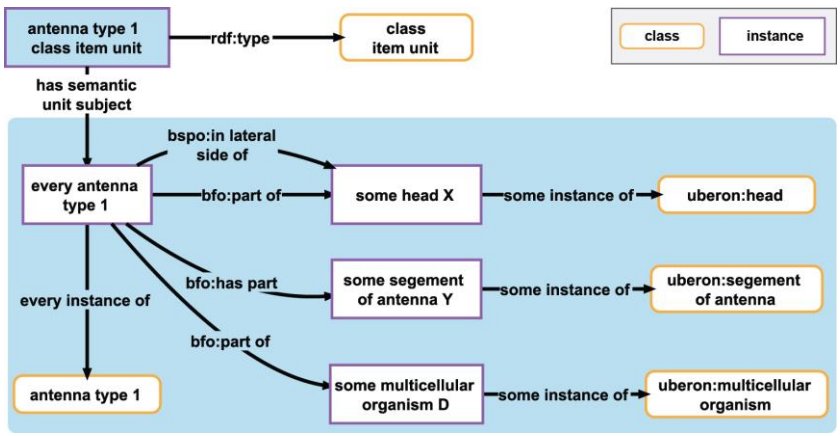


图12：类项单元。类项单元的一个示例（蓝色边框框出），其数据图显示在无边框的蓝色框中。该数据图由四个全称陈述单元的数据图合并而成，这些陈述单元的语义单元主体均为*「每个天线类型 1」*，并合并了与之对应的识别单元。

为使呈现内容清晰起见，未显示关联语义单元的资源和关系。

3. **文本-资源混合项单元**：当需要描述某项资源，而人们不愿意或无法使用三元组陈述和本体中的恰当术语，以机器可处理的形式化语义方式对其进行恰当描述时，就会使用文本-资源混合项单元。例如，在众包知识图谱中，用户想要描述某个具体的研究目标或某篇研究论文的主要贡献。该用户是领域专家，但很可能对语义学所知有限，甚至一无所知。在这种情况下，最好将研究目标或主要贡献指定为一项资源，该资源是以下项目的实例

或子类，例如目标规范（IAO:0000005），并为其提供文本描述，而不是用三元组对其进行形式化描述。文本描述既可以通过相应的注释属性记录为资源的描述，并以字符串作为其数值；也可以在 Neo4j 中记录为一个表示（IAO:0000219）该资源的数值节点。在知识图谱应用中，随后可以通过识别知识图谱已知的实体，自动为该文本描述添加语义注释。这些已识别的实体资源可以通过“提及”关系（IAO:0000142）与“目标规范”资源相连。由此形成的三元组集合构成一个表示文本-资源混合项单元的数据图。文本-资源混合项单元通过文本描述提供人类可读的信息，机器也可以在一定程度上利用“提及”关系处理这些信息。

项目组单元

项目组单元由至少两个项目单元组成（或由一个项目单元和一个主语不同的类型化陈述单元组成）。这些单元通过陈述单元彼此在语义上相关联，而这些陈述单元按照以下方式关联其主语：

如果 资源*X* = 是项目单元*1*的主语，资源*X* = 是陈述单元*1*的主语，资源*Y* = 是陈述单元*1*的宾语，且资源*Y* = 是项目单元*2*的主语，则陈述单元*1*在语义上连接项目单元*1*和项目单元*2*，构成一个项目组单元。

我们将以此处所述方式连接两个项目单元的陈述单元称为**项链接**。在语义单元图层中，这通过属性 **hasLinkedSemanticUnit** 将*itemUnit1*的资源与*itemUnit2*的资源关联起来，并通过属性 ***objectDescribedBySemanticUnit*** 将连接用的陈述单元与*itemUnit2*关联起来。尽管每个个体项链接都有特定方向，用于连接两个项目单元，但两个项目单元可以通过多个项链接相连，且这些链接的方向可能相反。例如，在描述一个包含所有成员的家庭时，描述父亲的项目单元会将父亲作为女儿的父亲，连接到描述女儿的项目单元；而描述女儿的项目单元则会将女儿作为父亲的女儿，连接到描述父亲的项目单元。因此，没有普遍适用的规则可以确定某个特定资源是项目组单元的主语，所以不指定主语。

除了项目单元外，项目组单元还可以由个体陈述单元、类型化陈述单元和质量测量单元组成；这些单元不属于任何特定项目单元，因为没有其他陈述单元与它们共享同一主语。

在大多数情况下，项目组单元会包含有关同一宾语的所有数据，其中一些项目单元包含与该宾语的部分或特定方面相关的数据。例如，某一特定学术出版物的所有数据都可以理解为代表一项研究活动的成果。与该研究活动相关的所有数据都可以通过一个项目组单元来组织和表示，其项目单元包含陈述单元，用于描述该活动的具体部分，例如更小的研究步骤、所采用的方法、材料、仪器以及参与该研究活动的研究主体。

由于每个项目单元中的信息都在用户界面中单独呈现在各自的页面上，人类读者可以在用户界面中以一组相互链接的页面的形式访问项目组单元。

可以将项目组单元分为实例条目组单元、类条目组单元和类公理条目组单元。**实例条目组单元**仅关联语义单元，其主语指向**命名个体**。另一方面，**类条目组单元**仅关联语义单元，其主语指向特定类的**某个实例**或**每个实例资源**。与类条目组单元关联的语义单元对特定类作出形式化描述，而该类又是类条目组单元的主语。如果类条目组单元的主语同时也是每个实例识别单元的主语，那么它就是**类公理项组单元**。类公理条目组单元是形式化的类描述，由于包含关系链，因此无法用单个类项单元来表达；它们还包含主语各不相同的全称陈述单元和偶然陈述单元，所以必须建模为项目组单元（图 13）。

如上文所述，在描述逻辑中（因此也在 OWL 中），全称陈述是在 TBox 中建模的；映射到 RDF 时，TBox 会包含空白节点的使用（例如，见图 5）。此类空白节点在三元组中起到类似匿名资源的作用（即 OWL2 中的匿名个体实例（OWL:AnonymousIndividual）），其使用在许多情境中都会造成问题，主要原因是：TBox 中以匿名资源表示的每个实体和属性，都~~无法在相应类公理之~~外被识别和指称。这会带来重大的实际影响，显著降低 TBox 表达式在经验数据建模方面的整体可用性。这些影响涉及以 TBox 建模的任何表达式的可查找性、可访问性、可探索性、可比较性、可扩展性、可复用性，以及整体机器可操作性。因此，经验数据最好建模为 ABox 表达式（详细讨论见（30））。在对类公理建模时，如果多个三元组需要指向同一个空白节点，空白节点就会造成问题

。OWL 所依据的描述逻辑采用一种树模型性质，将 OWL 公理限制为树形结构；因此，在使用 OWL 描述涉及三角关系的类公理时，常常会遇到困难（50）。

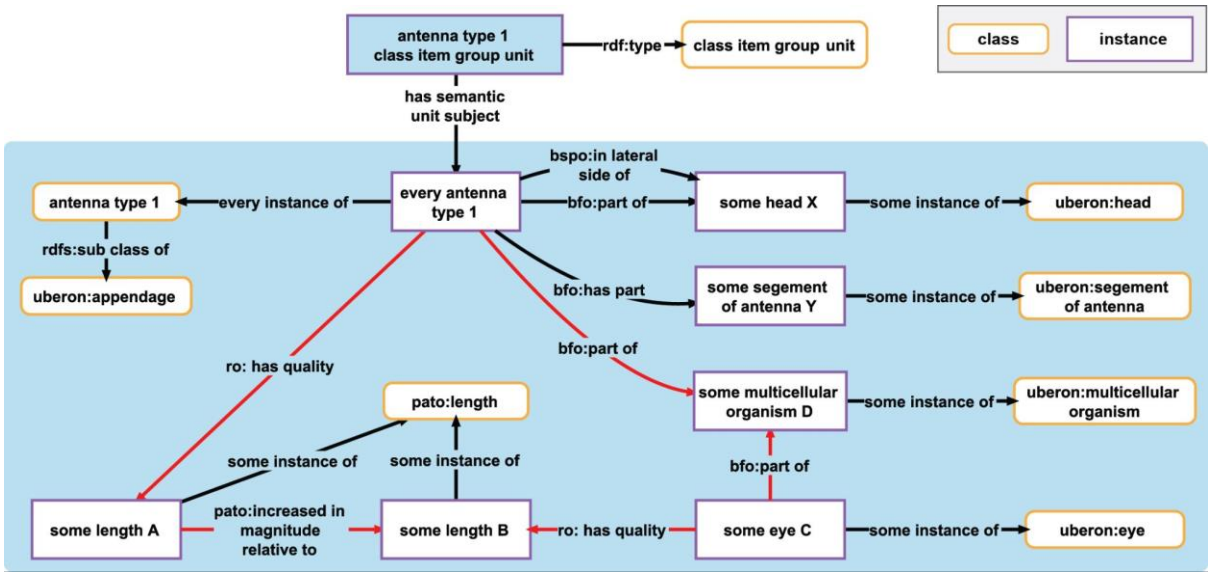


图 13：类公理项组单元。类公理项组单元的一个示例（带边框的蓝色框），其数据图显示在无边框的蓝色框中。该数据图由多个全称陈述单元及其对应的识别单元的数据图合并而成。这些语句无法用单个类项单元建模，因为它们并非都具有相同的主语。与 OWL 不同，这种表示法允许描述如图中红色所示的三元关系，表示每一个“天线类型 1”实例都比属于同一生物体的任意一只眼睛更长。为使展示更加清晰，未显示关联语义单元的资源和关系。

例如，如果你想正式描述一种特定类型的触角，它比其所属生物体的眼睛更长，可以尝试使用曼彻斯特语法 (51) 通过以下表达式对相应的*触角*类进行建模：

> ‘具有部分某个 ((触角 属于**某个多细胞生物**) 且具有质量某个 (长度且相对于某个（长度且内在于某个（眼 属于**某个多细胞生物**））相对于……增大))’。

然而，由于“某个多细胞生物”被提及的两处都是匿名的，因此该表达式未能体现预期含义：两处指的应是同一个多细胞生物。按照该表达式，只要某个触角比世界上某个多细胞生物的至少一只眼更长，就可以成为该类的实例；这只眼不一定属于拥有该触角的同一个生物体（52）。由于全称陈述单元不使用空白节点，不同陈述可以指向同一资源“某个多细胞生物 D”，因此不会遇到 OWL 的这一限制（见图 13，红色所示的三角关系）。

粒度树单元

由于每个陈述单元在知识图谱中都由其自身的资源（即节点）表示，并且会实例化相应的陈述单元本体类，而该类的特征由其实例所建模的命题所依据的关系类型决定，因此我们可以为该类添加更多信息。例如，我们可以识别依赖偏序关系的陈述单元类型。偏序关系是一种二元关系 *p*，具有传递性（如果 *A* 与 *B* 存在关系 *p*，且 *B* 与 *C* 存在关系*p*，那么 *A* 与 *C* 存在关系 *p*）、自反性（*A* 与自身存在关系 *p*）和反对称性（如果 *A* 与 *B* 存在关系*p*，且 *B* 与 *A* 存在关系 *p*，那么 *A* 与 *B* 相同）。此类偏序关系的例子包括全称陈述中的类-子类关系，或断言性陈述中的部分关系；此外还存在许多其他偏序关系，包括诸如在……之前（RO:0002083）这样的顺序关系，它们能够标识知识图谱中的**时间线**。

偏序关系可以产生粒度划分，进而形成**粒度树** (53–55)，并可用于定义**粒度视角** (56–58)。粒度视角能够识别知识图谱数据图层中具有语义意义的特定类型树状子图，并可用于支持对知识图谱的探索，其方式与陈述单元、项目单元和项目组单元有一定的正交性。通过识别可能与粒度视角相关的陈述单元类（因为其特征关系属于偏序关系），应用程序可以利用这些信息，在知识图谱中自动识别相应类型的粒度树，并使其可供用户访问。

由于粒度树具有嵌套结构，且其方向性从根节点隐含地指向叶节点，因此可以将**粒度树单元的主语**指定为满足以下条件的陈述单元的主语：这些陈述单元的宾语与主语共享，而它们的主语不与属于该粒度树单元的其他陈述单元的宾语共享。在语义单元图中表示粒度树单元的资源，通过属性 **hasSemanticUnitSubject** 与其主语关联。

粒度树单元的特点是，其陈述单元始终是同一陈述单元类的实例（因为该类是根据底层偏序关系来定义的），且主语和宾语始终是同一种基本类型的资源。为了使粒度树单元不仅可由机器处理，也更易于人类理解，它们包含与构成粒度树的陈述单元相对应的类型化陈述单元。

由于语义单元图中表示粒度树单元的资源，通过 **hasAssociatedSemanticUnit** 属性与其关联的陈述单元相关联，因此只需查询这些关系，就能识别并区分不同的粒度视角

以及通过 *hasSemanticUnitSubject* 关系与粒度树单元的主语相关联，从而无需进一步将粒度树单元的类划分为子类。

由特定类型的陈述单元衍生出的粒度视角，为将整个数据图划分并组织成语义上有意义的较小数据图提供了另一种方式，因此也构成了另一类复合单元。这类粒度视角的例子包括具有部分关系构成的树状层级，例如生物体内各种解剖结构之间的关系；特定过程中的事件时间序列；发育依赖链；更广泛的因果关系链；以及基于类-子类关系的概念分类体系。

粒度项目组单元

粒度项目组单元由主体属于同一粒度树单元的所有陈述单元、质量测量单元和项目单元组成。粒度树单元的主语也就是**粒度项目组单元的主语**，而语义单元图中表示粒度项目组单元的资源通过属性 *hasSemanticUnitSubject* 与其主语相关联。

粒度树为相关资源提供嵌套层次组织。如果这些资源是项目单元的主体，则相应的项目单元可以按照粒度树提供的层级进行组织，从而将多个项目单元编排成嵌套层次结构。知识图谱应用可以利用对数据图层的这种额外组织和结构化，提升用户对整个数据图的**可探索性**。因此，粒度树单元可以通过将项目组单元中的项目单元组织成粒度项目组单元，为项目组单元提供额外的组织结构。当知识图谱高度互联、因而其中的项目组单元非常庞大时，这一点尤为重要。

由于语义单元图中表示粒度化项目组单元的资源通过属性 *hasAssociatedSemanticUnit* 与其关联语义单元建立关系，因此，只需查询这些关系以及它们各自主体的 *hasSemanticUnitSubject* 关系，即可识别不同类型的粒度化项目组单元，无需进一步将粒度化项目组单元的类细分为子类。

上下文单元

上下文单元由所有语义单元组成：合并这些单元的数据图后，会形成一个连通图，其中没有任何资源或三元组与其他资源或三元组分离；也就是说，任意两个

上下文单元的数据图中的资源通过一系列同样属于该上下文单元数据图的三元组语句彼此相连。此规则唯一的例外是属性为 **关于陈述单元**的三元组，其属性为 关于（IAO:0000136）。该属性将信息制品与制品所包含信息所涉及的实体关联起来。因此，它总是通过将语义单元资源设为主语、将数据图中的某个其他非语义单元资源设为宾语，来标记从话语层到本体层的参照框架变化（有关层的讨论，参见 关于陈述的陈述：使用语义单元在知识图谱中记录本体、诊断和论述信息）。换言之，**关于陈述单元将语义单元图中的资源与知识图谱数据图中的资源关联起来，可用作索引，标示图中参照框架发生变化的位置**。例如，记录一项研究活动及其输入和输出，而输出是一个数据集，其中包含对某个多细胞生物解剖结构的描述。图中将包含陈述 *'描述条目单元'* 关于'多细胞生物'（UBERON:0000468），它指示参照框架从研究活动的结果（即描述）转变为描述所涉及的多细胞生物。简而言之，它表示从研究活动参照框架转变为研究对象参照框架（另见下文图 14）。因此，关于陈述单元总是标示属于两个不同上下文单元的数据图之间的边界。与项目组单元类似，由于缺乏普遍适用的规则来识别作为项目组单元主语的特定资源，因此不指定主语。

数据集单元

数据集单元是由特定语义单元构成的、有明确定义且经过排序的集合。这个集合可以是同质的，即只包含同一类型的语义单元；也可以由各种陈述单元和复合单元汇编而成。与其他语义单元一样，每个数据集单元在语义单元图中都有自己的节点和 UPRI。

数据集单元有多种用途。例如，可以汇集合作项目中特定机构贡献的所有数据，也可以记录特定宾语在特定时间的状态。此外，特定搜索查询的结果也可以存储为数据集单元并供人访问。知识图谱的用户可以指定供自己使用的特定数据集单元，并将该单元的 UPRI 用作标识符来指称它。

除了将数据集明显地用作特定语义单元的集合之外，数据集单元还可用于管理关于陈述的陈述：如果用户想指出，来自某一出版物（项目组单元 *X*）的特定测量数据（断言陈述单元 *A* ）

支持或直接反驳了另一出版物（项目组单元 *Y*）中的某个因果理论（全称陈述单元 *B*），他们可以创建相应的断言陈述单元 *C*，将陈述单元 *A*与支持性陈述单元 *B*关联起来；也可以创建一个由这三个陈述单元 *A*、*B*和*C*组成的数据集，而不包含*X*和*Y*。虽然在用户界面中，这可能不会被呈现为数据集，但从技术上讲，它可以作为数据集单元进行构建、存储和管理。

列表单元

有时，您需要对特定资源列表作出陈述。为此，可以将这样的列表建模为一种特定类型的复合单元，即列表单元。列表单元的使用方式与其他类型的语义单元相同：每个列表单元都是一个语义单元，在语义单元图中由其自身的节点表示，拥有自己的 UPRI，并实例化相应的列表单元本体类。

我们区分**无序列表单元**和**有序列表单元**，后者中的资源按特定顺序排列，例如学术出版物的作者。另一方面，**集合单元**是一种无序列表单元，其中每个资源只列出一次（唯一性限制）。

从技术上讲，列表单元是由多个成员关系陈述单元构成的集合，每个成员关系陈述单元都通过一条 *子项* 关系，将列表单元的 UPRI 与相应资源连接起来，从而指定列表中的一个资源。对于有序列表单元，必须先为每个成员关系陈述单元编制索引（通过数据属性索引 (RDF:index)，将成员关系陈述单元的 UPRI 与从 0 开始连续递增的整数值连接起来），从而形成带索引的成员关系复合单元，然后再将其与列表单元关联。

列表单元可用作数组，并且可能包含基数限制，从而描述实体的封闭集合，形成局部封闭世界假设。

讨论

作为FAIR数字对象的元数据和语义单元

欧盟委员会FAIR数据专家组提出了**FAIR数字对象**这一概念，并认为它是实现 FAIR 的核心（16）。FAIR数字对象必须配备持久标识符、元数据和上下文文档，以便可靠地发现、引用和重复使用。属于 FAIR数字对象的数据应以通用且理想情况下

采用开放文件格式存储，并使用元数据标准和成熟的词汇表进行详尽记录。

FAIR数字对象也代表着欧洲开放科学云（EOSC）互操作性框架的核心概念：它们是FAIR生态系统中的原子实体，通过提供实现以下五个互操作性层所需的元数据发挥作用：1）信息技术（IT）系统在实现或访问时，必须能够与其他IT系统协同工作，且不受任何限制，或采用受控访问方式，以实现技术互操作性；2）通用数据格式和通信协议，以实现语法互操作性；3）与通用语义资源相关的上下文语义，以实现语义互操作性；4）与通用流程资源相关的上下文流程，以实现组织互操作性；5）与通用许可资源相关的上下文许可，以实现法律互操作性（15）。相应的元数据本身也可构成FAIR数字对象，并可通过其持久标识符相互链接。因此，FAIR数字对象可以具有不同的粒度层级；较粗粒度层级的FAIR数字对象可能由若干较细粒度层级的FAIR数字对象组成。

语义单元的概念可以调整为适用于FAIR数字对象。每个语义单元都有自己的UPRI，因此已经具备所需的持久标识符。此外，还可以通过网页界面使用SPARQL和CYPHER等通用规程访问和搜索这些单元；网页界面采用HTML，并支持RDF、JSON及其他格式作为数据导出格式。如果知识图谱采用成熟的受控词汇表和本体术语，并使用SHACL（32）、ShEx（59,60）或OTTR（61,62）等工具，按照既定图模式组织数据和元数据，那么语义单元的数据图中的所有数据都将具备语义互操作性。此外，每个语义单元都可以单独指定版权许可，该许可既可以由知识图谱应用自动提供，也可以由特定语义单元的创建者单独指定。

通过添加三元组，可以采用相应的属性—值对（属性图）或使用成熟的来源词汇表创建额外的来源命名图（RDF/OWL），以追踪每个特定语义单元的相关来源数据。**语义单元的溯源元数据**可以涵盖创建者、创建日期、创建应用程序、语义单元的标题、所有贡献用户以及最后更新日期等信息，并且仅限于有关语义单元本身的来源数据——它们不涉及数据生产的原始过程，因此也不涉及该单元数据内容的溯源。

此外，如果知识图谱应用区分已发布和未发布的数据，则可以单独跟踪**出版元数据**，包括语义单元的发布日期、创建者及其贡献者名单。

访问控制元数据指定任何版权许可；如果知识图谱应用区分不同的用户群体或角色，还会说明哪些人有权访问该语义单元。

关于数据的 FAIR 原则，建议同时提供**建模元数据**，即有关在该单元的数据图中用于对数据进行建模的数据模式或图模式的信息。给定语义单元类的每个实例最好都应用相同的图模式。为此，最好能将各自具有独立 UPRI 的模式和图模式存储在某个开放存储库中，或将这些模式和图模式记录在相应的语义单元本体类中。

指向数据生产过程的溯源元数据（参见例如图 4）可以采用元数据陈述单元以及元数据项和项目组单元的形式进行记录。数据生产的溯源通常比该单元自身的溯源更加详细和复杂，必须将其记录为语义单元，从而使其本身成为 FAIR数字对象。

实现

语义单元、纳米出版物和微型出版物

作为 FAIR数字对象的断言式陈述单元这一概念与纳米出版物非常相似——尤其是在 RDF/OWL 中实现时。**纳米出版物**是 RDF 图，表示从文献中提取并以溯源和归属信息丰富的最小可出版信息单元（63–66）。它们利用命名图和语义网技术。每个纳米出版物都以机器可读的格式和语义对某个特定断言（例如科学主张）进行建模，并可通过唯一标识符访问和引用。纳米出版物用于促进学术断言的发现、探索和再利用（66）。此外，嵌入采用语义模板的去中心化网络后，纳米出版物还可供用户直接发布小型关联数据陈述，并将其作为纳米出版物发布（65）。

每个纳米出版物由四个命名图组成：

1. 头部命名图，将其他三个命名图连接到该纳米出版物的唯一标识符；

2. 包含以图形式建模的断言的断言命名图；

3. 包含有关该断言的元数据的溯源命名图；以及

4. 包含有关纳米出版物本身元数据的出版物信息 命名图 If.

纳米出版物非常适合用于表示和发布断言陈述单元，其中断言 命名图与断言陈述单元的数据图密切对应，而头部 命名图则与其语义单元图对应。溯源命名图中的三元组可能链接到其他语义单元，进而链接到包含详细元数据描述的其他纳米出版物（例如图 4 所示的图）。纳米出版物也适合按照同一模式发布其他类型的陈述单元。另一方面，复合单元对应于带注释的纳米出版物集合，这些集合既可以静态定义（例如纳米出版物索引（67）），也可以动态定义（通过在知识图谱上指定查询）。

通过使用数据集单元，可以将多个此类语义单元纳米出版物组合起来，形成一篇微型出版物。**微型出版物** （68）用于表示科学论证，其中一项特定断言代表一项主张，而其他断言则以数据及其附带元数据的形式提供支持或反驳该主张的科学证据，同时还包括归属信息、支持性和反驳性参考文献、指向经验数据和图表的链接等信息。还可以添加更多断言，指出与其他科学论证所提出的主张之间的关系。通过这种方式，微型出版物将以纳米出版物形式呈现的科学主张与所有相关信息联系起来，使主张构成科学论证（*科学论证*可以构成一种特定类型的数据集单元）。

使用纳米出版物模式将知识图谱组织成相互关联的语义单元

将知识图谱组织成相互关联的数据图（每个数据图都作为特定语义单元组织起来）的第一步，是在三元组层级之上创建一个基础抽象层。实现方法是将知识图谱划分为一组陈述单元，即建立三元组到陈述单元的映射，其中每个三元组恰好属于一个陈述单元的数据图。

在 RDF/OWL 中实现时，可以将陈述单元建模为纳米出版物（见上文）。在 Neo4j 中，由于它是属性图，要识别属于同一陈述单元的所有三元组，必须为该单元的每个节点和关系添加一个‘陈述_单元_*URI:upri*’属性—值对，并将该陈述单元的 UPRI 作为其数值。

基于这一陈述单元的初级抽象层，可以组织出由复合单元构成的次级抽象层。每个复合单元都被定义为由两个或更多

语义单元（即陈述单元和其他复合单元）的集合，方法是将复合单元的 UPRI 与其关联语义单元的 UPRI 关联起来。在 RDF/OWL 中，可以将复合单元的语义单元图用作相应纳米出版物的 头部 命名图，同时将纳米出版物的断言 命名图留空，以此实现。因此，头部 命名图指定了与该复合单元关联的所有基本单元和复合单元。相比之下，在 Neo4j 中，属于某个复合单元的所有三元组的节点和关系都具有一个值为该复合单元 UPRI 的'复合_单元_*URI:upri*'属性—值对。由于一个给定的陈述单元可以关联多个复合单元，其'复合_单元_*URI*'属性的值可以是不同语义单元的多个 UPRI 组成的数组。

模块化提高了数据管理的灵活性

> "人类思想具有组合性（简单的部分可以组合）和递归性（部分可以嵌入其他部分），因此只需有限的思维工具，就能探索广阔无垠的知识。"（（69），第360页）。

与人类思想类似，将知识图谱构建为彼此独立但具有递归结构的子图，每个子图都属于某个特定的语义单元，而这些语义单元又被划分为不同的语义单元类，这样便将知识图谱组织成**结构化模块**，这些模块可以**（递归地）组合**，并能**将复杂性封装为易于管理的单元**，从而显著提高图的表达能力和数据管理的灵活性。通过允许指定语义单元之间的关系，较细表征粒度中的某个语义单元可以关联到较粗粒度中的多个单位。因此，某个陈述单元可以关联到多个复合单元，同时，该陈述单元本身及其所有三元组都记录在图中的同一个位置。

例如，多细胞生物的描述可能包含一个具有部分关系陈述单元，说明该生物体具有头部作为其部分。这个陈述单元将与该生物体自身的项目单元相关联，而该项目单元又会与描述该生物体其他部分的更多项目单元相连接，从而构成一个项目组单元。此外，由于具有部分是一种偏序关系（70），这个具有部分关系陈述单元还将与一个部分关系粒度树单元及其对应的粒度项目组单元相关联。因此，这个具有部分关系陈述单元将至少与另外四个语义单元相关联。

由于语义单元采用模块化组织，实现这些语义单元的知识图谱可以构建为**虚拟联邦知识图谱**，其中的数据由多个项目和机构贡献，并存储在各自的存储库中。这样，能够访问所有数据的用户就可以利用该知识图谱的

内容，同时数据管理仍由各参与贡献的项目和机构掌握，从而确保其技术自主权。

语义单元带来的内容模块化，也有助于基于分区查询知识图谱。其他图分区方法已成功实现这一点（71）；我们预计，将图划分为语义单元，从而形成语义上有意义的子图，对查询任务也会产生类似的效果。然而，语义单元在知识图谱分区和模块化中的应用仍有待未来研究。

利用语义单元将知识图谱组织为五个表征粒度层级、其他粒度树以及不同的参照框架

以陈述单元、项目单元和项目组单元作为语义单元的不同类别，知识图谱可以组织为以下基本表征粒度层级（图 14）：三元组、陈述单元、项目单元、项目组单元，以及作为整体的知识图谱。三元组代表知识图谱中可能的最低、也就是最精细的抽象层级；语义单元则构成更粗略的抽象层级，因此也构成更粗略的表征粒度层级。**表征粒度将知识图谱的语义单元图层，也就是话语层，组织为不同的表征粒度层级**，并由此间接组织知识图谱的数据图层。

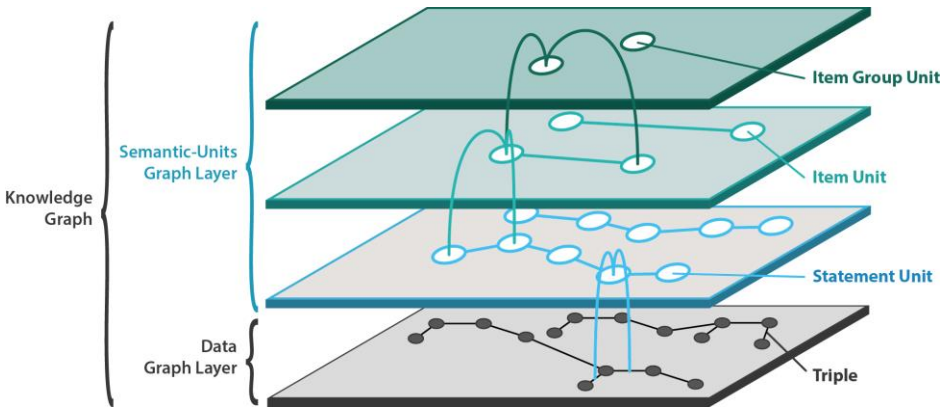


图14：五个表征粒度层级。 The 语义的引入 单位至 一个知识 图谱会增加一个语义单元- 单位图层 数据图层，它 还增加了 陈述单元层级 ，a 项目单元层级 ，以及一个 从项目组单元层级到三元组层级，再到整个图的层级，由此形成五个表征粒度层级。

将三元组组织成陈述单元（即对人类读者而言，代表语义上有意义的最小命题单位），再将陈述单元组织成项目单元（可呈现为

用户界面中的独立页面），并将项目单元组织成项目组单元（可呈现为语义上相互关联的用户界面页面集合），通常是提升知识图谱的人类可读性和可用性的必要做法。

除了将知识图谱组织和结构化为不同层次的表征粒度之外，还可以利用项目组单元中存在的粒度树单元，以一种大致正交的方式，将给定的项目组单元进一步组织和结构化为粒度化项目组单元。**粒度树将知识图谱的数据图层以及本体层组织为不同的粒度视角**。

最后，还可以将给定的项目组单元进一步组织和结构化为不同的上下文单元，每个上下文单元都涵盖特定的参照框架。不同参照框架之间，因而不同上下文单元之间的接触点，可以通过关于关系陈述加以识别：关于关系陈述的主语是一个语义单元资源，其主语属于某个上下文单元的数据图，而关于关系陈述的宾语属于另一个上下文单元的数据图（另见图16）。这样，便可以通过知识图谱中关于关系陈述的连接，在不同参照框架之间进行遍历。**上下文单元将知识图谱的数据图层以及本体层组织为不同的参照框架**。

将数据组织为各种语义单元，从而构成具有语义意义的子图，使人类读者能够将其作为语义相关的较大数据块来理解，是在用户界面中呈现数据的有效方式。通过将知识图谱细粒度地组织成不同类型的语义单元（其中一些属于不同的表征粒度层级），并将知识图谱构建为这些部分重叠、部分包含的子图，可以显著改善数据管理、数据访问、数据探索和图导航。

使用语义单元表达关于陈述的陈述，并在知识图谱中记录本体、诊断和论述信息

由于每个语义单元都有自己的 UPRI，并且在知识图谱的语义单元图层中由各自的节点表示，因此很容易对**语义单元作出断言**，进而作出**关于陈述的陈述**。如果将语义单元组织为 FAIR数字对象，还可以跨不同的数据库和知识图谱作出陈述。通过将知识图谱构建为语义单元，可以开发用户界面，提供直观易用的框架，用于

作出关于陈述的陈述。开发此类用户界面及其配套工具和服务时，若语义单元按照清晰、简明的模式实现，便能从中受益。这样，语义单元有助于解决在知识图谱中作出关于陈述的陈述可能颇具挑战这一问题（见挑战*3*）。知识管理和知识表示中还有另一个方面：以简便的方式对陈述作出陈述，并拥有一种用于表示偶然陈述的形式化记法，都会带来益处。

作为一名科学实在论者，我们可以区分：（1）作为独立于心灵的**真实实体**的个别事物和共相；（2）我们以观念、概念和思想的形式对**真实实体的认知表征**；以及（3）以词语、符号、图像、录音等形式对**真实实体的文本表征和基于感知的表征**（72–74）。后者通常用于交流和记录，发送者意在触发接收者的认知表征，使其与发送者自身的表征尽可能相似。

意义和**指称**是交流成功的关键，并决定了文本表征的双重性：文本表征必须同时提供本体信息和诊断信息。所传递的信息必须具有**语义意义**，也就是说，它能够回答“它是什么？*</style”*这类问题，并为所有被指称的实体提供承载语义概念内容的本体定义。其中一部分本体信息会明确提供，但在大多数情况下，发送者会通过使用公认术语，依赖接收者的推理性词汇能力（75）。

但交流也必须能够回答“它看起来是什么样？”这类问题，从而提供有关认识论表象的信息——这些信息对于**明确指称**文中提及的所有实体至关重要。这类**诊断**信息，即**指称信息**，在接收者已经认识该实体时或许并不重要；但在其他所有情况下，尤其是涉及理论实体时，其存在无法被直接识别或观察，且往往无法从本体定义中推导出识别标准（例如铅原子、细胞核），诊断信息对于可靠指称至关重要。因此，需要明确提供部分诊断信息，但在大多数情况下，发送者依赖接收者的指称性词汇能力（75）。

这与知识图谱有什么关系？有时我们谈论真实实体，有时则谈论对这些实体的认知表征和文本表征。有时，这些讨论会相互重叠，而在知识图谱中对其进行建模时，复杂性便由此而生。理想情况下，本体定义和诊断性操作定义由

领域本体提供，知识图谱中的断言则指向相应的本体术语及与之关联的信息。这样，知识图谱便能高效地记录、表征和传达经验数据。

然而遗憾的是，由于 OWL 和描述逻辑在定义类方面存在局限，本体通常不提供诊断性操作定义。本体以亚里士多德定义（76）的形式定义类，这些定义构成本质主义类。**本质主义类**以**全称陈述**为基础，类成员资格取决于一组**各自必要**且**合起来充分**的属性（77）。然而，许多诊断性操作定义提供了依赖于方法的识别标准，要求对宾语进行操作才能确定其类成员资格，例如宾语小到肉眼无法看见，或者只有在染色后或使用某种检测试剂盒时才能识别。此类操作性定义的界定属性往往指的是并非每一个实例都必然具有的属性。相反，这些属性通常只是典型特征，且只有达到特定数量时才足以作为判断依据。因此，许多识别标准无法建模为本质主义类，而必须建模为簇类（74）。与本质主义类不同，**簇类**的实例不必具备所有界定该类的属性——只需其中一些属性适用，而充分条件则是达到规定的百分比阈值，即**最低配额**（77）。因此，簇类的定义采用**偶然陈述**的形式，无法在 OWL 和描述逻辑中表示（见上文讨论）。

诊断性操作定义往往不仅依赖簇类概念，也依赖模糊集，因为许多簇类定义都依赖于指称**模糊集**的术语（74）。例如，当识别标准使用粗糙的内部物质、电子致密簇、蘑菇状、卵圆形或钩状等术语，来识别某一特定解剖结构类型中实例的类成员资格时，这些术语指称的是用于描述时空结构属性的词汇；与文本定义相比，通过范例或图像，也就是感知内容，能更好地传达这些属性。虽然本质主义类和簇类始终以文本定义，但模糊集可以通过指称范例来进行指示性定义，从而指向基于感知的内容；这类内容依赖人类的模式识别能力。

语义单元方法为表征偶然陈述提供了一种形式化记法；如果再结合经语义标注的示例图像，知识图谱便能提供回答“看起来怎么样？”这类问题所需的全部诊断知识。结合本体定义，知识图谱还可向任何读者提供其所需的全部相关知识，以确保其具备推理性和指称性词汇能力。

除了本体层和诊断层之外，知识图谱通常还会记录第三层信息，即**论述语境信息**（另见 Ingvar Johannson 对语言实体使用与提及的区分（78））。换言之，知识图谱不仅包含以断言形式呈现的经验数据，例如“铅的熔点为 *327.5 °C*”，还包含诸如“作者 *A* 断言铅的熔点为 *327.5 °C*”或“关于铅的熔点为 327.5 °C 的断言，是实验 X 的结果”之类的断言， 这些断言涉及论述语境信息。有些断言甚至会涉及诊断层面，说明文本表征与真实实体之间的指称如何实现，例如记录相应的方法和诊断标准。若没有一种有效的方法来表述关于陈述的陈述，就无法记录和表征这三层信息（即**本体**、**诊断**、**论述**），也无法表达它们之间的相互关系。

由于语义单元在图中以各自独立的节点表示，因此能够支持对这三个层及其相互关系的记录与表示。陈述单元可以在其数据图中提供**本体信息**，这些信息可以汇集到表征粒度较粗的复合单元中。作为命题，它们在语义单元图中表示，构成**话语层**的重要部分。此外，与 OWL 和描述逻辑不同，语义单元方法通过引入某个实例和每个实例资源，以及用于表示个别全称陈述和语义相关的全称陈述集合的资源，使人们能够对与本体层相关的各种实体进行细致的指称。因此，人们可以对个别类（=概念）、它们的定义、各自的定义性陈述、某个类的单个实例、某个类的实例组，以及某个类的所有实例组成的组进行陈述，从而为讨论全称陈述及其组成元素开辟多种途径。此外，借助上下文单元，甚至可以区分并明确记录本体层和话语层中的**不同参照框架**。换言之，语义单元方法能够指定其知识图谱的话语层与本体层之间的各种相互关系，因此将全称陈述纳入知识图谱的论域（挑战*4*）。

由于陈述单元还可以将特定资源与另一个陈述单元关联起来（例如，‘作者_-断言-> 陈述_单元_*Y*’），或将两个陈述单元彼此关联起来（例如，‘陈述_单元_*X* -具有元数据-> 陈述_单元_*Z*’），因此很容易记录这三个层次之间的关系。例如，项目组单元记录了某一特定学术出版物的内容。由于学术出版物通常可以被视为关于某项科学调查或其他研究活动的报告，因此项目组单元将包含有关某项研究活动的信息

研究活动是一种过程，可能以某些物质实体（即物质、仪器、标本等）和研究主体（即实验者、患者、受试者等）或某些数据为输入，并以某项研究结果为输出。此外，研究活动通常会实现一项特定计划（即研究方法），并达成特定的研究目标，而研究目标是计划的一部分（见图15）。这一通用模型可以像俄罗斯套娃一样应用：研究活动可以将其他研究活动作为其部分，即整体研究过程中的不同步骤；这些步骤也各有自己的输入和输出，实现各自的研究方法并达成各自的目标。

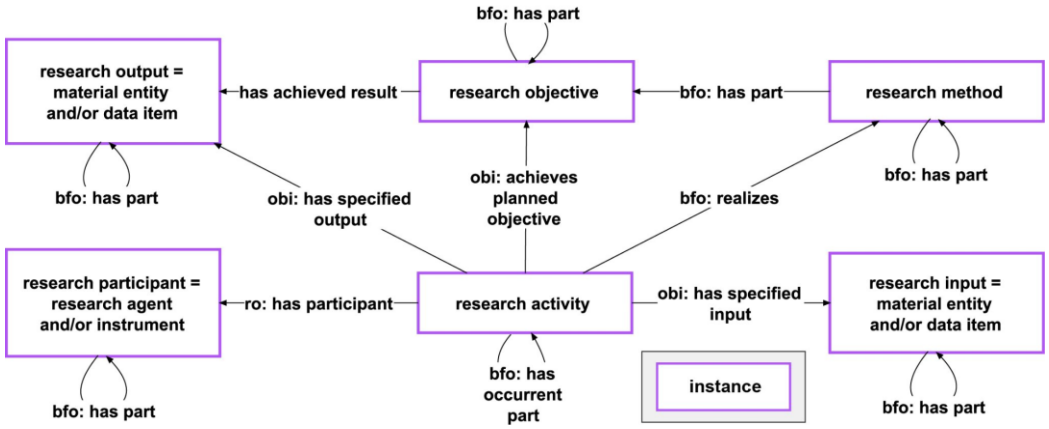


图15：用于建模学术出版物内容的数据模式。研究活动与其输入、输出，以及以研究方法和研究目标形式呈现的底层过程计划规范之间的关系（模型改编自（79））。

该模型可应用于学术出版物的知识图谱，并可用于构建语义单元。出版物的项目组单元关于一个*研究活动*实例，并关联一个以该*研究活动*实例为主语的研究活动项目单元。研究活动项目单元关联一个研究结果条目单元，该单元以一个*研究结果*实例为主语。研究活动资源通过属性 *hasSpecifiedOutput* (OBI:0000299) 与研究结果资源相连。现在，研究结果条目单元可以关联一个描述条目单元，其中包含若干关于某个特定多细胞生物的断言陈述单元——该出版物描述了某个特定多细胞生物的解剖结构。这通过一个三元组记录下来，该三元组通过属性 *hasLinkedSemanticUnit* 将*研究结果条目单元*与*描述条目单元*关联起来。'描述' (SIO:000136) 资源是*描述条目单元*的主语，并通过属性 *hasPart* (BFO:0000051) 与多细胞生物条目单元相关联；多细胞生物条目单元则通过属性 *hasSemanticUnitSubject* 与多细胞生物实例 (UBERON:0000468) 相关联。由此产生的语义单元集合涵盖至少三个不同的参照框架：出版物本身、出版物所报告的研究活动，

以及关于研究对象（即多细胞生物）的断言性陈述，其边界由两个关于关系陈述标示。图16展示了由此产生的话语层和本体层，以及不同参照框架之间如何相互关联。

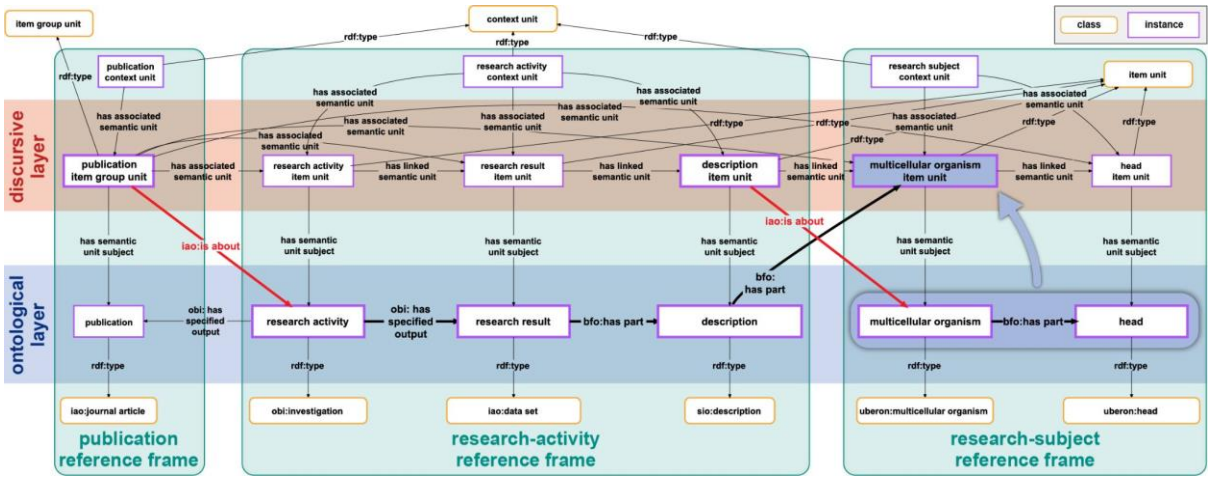


图16：D　学术出版物内容的项目组单元所属图的细节，呈现于 ABox RDF图中。所有内容均依据一种特定的数据模式（见图15）建模，该模式已针对语义单元进行了调整。已发表期刊文章中的所有内容都通过各种关联语义单元，组织在各自的项目组单元实例中。出版物本身在数据图中表示为期刊文章（IAO:0000013）的一个实例。该项目组单元包含若干与之关联的研究活动项目单元，而这些项目单元又通过属性 **hasLinkedSemanticUnit*彼此相连。描述研究活动的项目单元以调查（OBI:0000066）的一个实例为主语；该实例的输出是数据集（IAO:0000100）的一个实例，而该数据集以描述（SIO:000136）的一个实例作为其部分。该描述实例以描述多细胞生物的项目单元为其部分；后者以多细胞生物（UBERON:0000468）的一个实例为主语。蓝色箭头表示，数据图（无边框的深蓝色方框）由该项目单元（同色带边框方框）表示。该项目单元还关联着一个头部项目单元，其主语是头部（UBERON:0000033）的一个实例，而头部是多细胞生物的一部分。语义单元的数据图构成本体层，而其语义单元图则构成话语层。三个不同的上下文单元将出版物、研究活动和研究对象各自的指称框架区分开来，并以关于关系陈述标示其边界。为使呈现更清晰，话语层中未显示关联的陈述单元。

语义单元的资源表示并组织话语层，而其数据图表示本体层（以及诊断层）。这同样适用于所有语义单元资源。将知识图谱组织为这些不同的层以及不同的参照框架，可能为开发创新的可视化和图探索方法提供一种新的框架。

否定建模、基数限制与分歧 t

基于 OWL 和描述逻辑的信息系统遵循所谓的 **开放世界假设**（OWA）。OWA 默认假设信息是不完整的。因此，缺少关于某个实体或某个事实的信息，并不一定意味着该实体不存在，或意味着对该事实的否定。因此，例如，如果你想陈述某个特定宾语是果实但不是梨果、某个昆虫的特定头部不具有任何触角，或者该头部恰好具有三只眼睛，仅仅说明该宾语是果实、该头部具有三只眼睛，或不提及任何触角，都是不够的，因为根据 OWA，该果实仍可能是梨果，该头部仍可能具有触角，也仍可能具有第四只眼睛。相反，你必须明确陈述所有这些内容。

遗憾的是，在 RDF/OWL 中，这些类型的否定和基数语句会带来棘手的建模挑战，因为它们无法直接表示为实例之间的关系，因此也无法表示为 ABox 表达式。相反，必须将它们建模为 TBox 中的类表达式，例如使用 OWL 曼彻斯特语法 (51)。

关于**否定**，可以区分两种不同类型的否定语句：涉及类指称的否定，以及对实例间关系的否定。前者的一个例子是语句“这个果实不是梨果”（图 17A）。按照曼彻斯特语法，可以将其表述为“not (type pome fruit)”（图 17B）；转换为 OWL 并映射到 RDF 后，会得到一幅图，其中给定的特定实体是类果实 (PO:0009001) 的一个实例，同时也是类梨果 (PO:0030110) 的补类的一个实例（图 17C）。类似的情况还包括作为类公理一部分的否定。

语义单元可以通过引入一个通用的*否定单元*类，提供另一种建模方案（在 Neo4j 中，只需为相应的语义单元节点添加标签‘*:Negation*’，即可表示该资源是‘否定’类的实例）。陈述‘这种果实不是梨果’可以用两个语义单元建模，二者都是*命名个体识别单元*和*断言陈述单元*的实例。不过，陈述其主语是梨果（PO:0030110）的语义单元也属于*否定单元*的实例，因此会否定其数据图的内容（见图17D）。

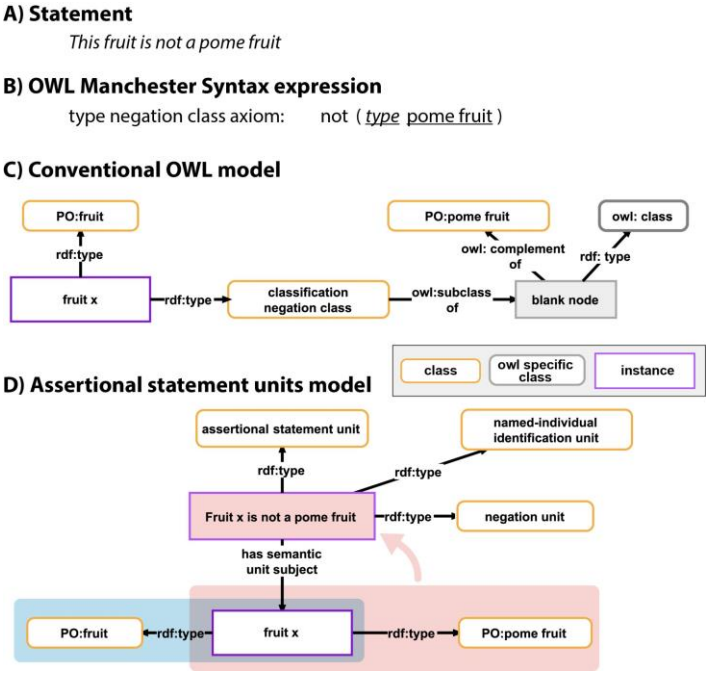


图17：运用语义单元对涉及实例的否定进行建模。 **A)** 人类可读陈述：这种果实不是梨果。该陈述可以通过两种不同方式建模。一种是 OWL 表达式，可使用**B)** 曼彻斯特语法指定，其中该果实是某个类的实例，而该类被定义为其所有实例都不是梨果（PO:0030110）的实例。请注意，这个曼彻斯特语法表达式如何转换为**C)** 映射到 RDF 的 OWL 表达式：果实 x 是果实（PO:0009001）的实例，同时也是梨果（PO:0030110）的补类的实例。~~或者，也~~可以使用语义单元将该陈述建模为基于实例的图**D)**。蓝色方框中的数据图表明该实体是果实。属于该数据图的

表示否定的陈述单元（带边框的红框，此处显示其动态标签）则显示在没有边框的红框中，并指出该实体是梨果。然而，由于后一个陈述单元不仅实例化了*断言陈述单元*和*命名个体识别单元*这两个类，也实例化了*否定单元*，因此它实际上否定了数据图中的陈述，从而表明它不是梨果。

缺失陈述是另一种涉及指称类的否定示例。观测"这个头部没有触角"可转换为这样的陈述：该头部是某个类的实例；按照曼彻斯特表达式，该类的特征为"not ('具有部分' some 触角)"。将其转换为 OWL 后，会得到一个更复杂的类公理（见图18A-C）。

同一语句可以建模为两个语义单元的子图的并集[6]，其中以"头部 x"（UBERON:0000033）为主语的子图是*具有部分关系陈述单元*、*断言陈述单元*和*否定单元*的实例，而以"某触角"（UBERON:0000972）为主语的子图则仅是*某实例识别单元*的实例~~（见图18D）。一般而言，以这种方式建模缺失陈述时，断言中所有旨在表示缺失实体类型的实例资源，都会按照这种方式建模，并通过其对应的某实例识别单元中的属性*某实例~~属于*与相应的类建立关系。与使用更复杂类表达式的基于 OWL 的表示法相比，这种表示法在知识图谱应用中更简单、更易于使用和实现。

[6] 子图的并集，但在语义上是语句的交集。

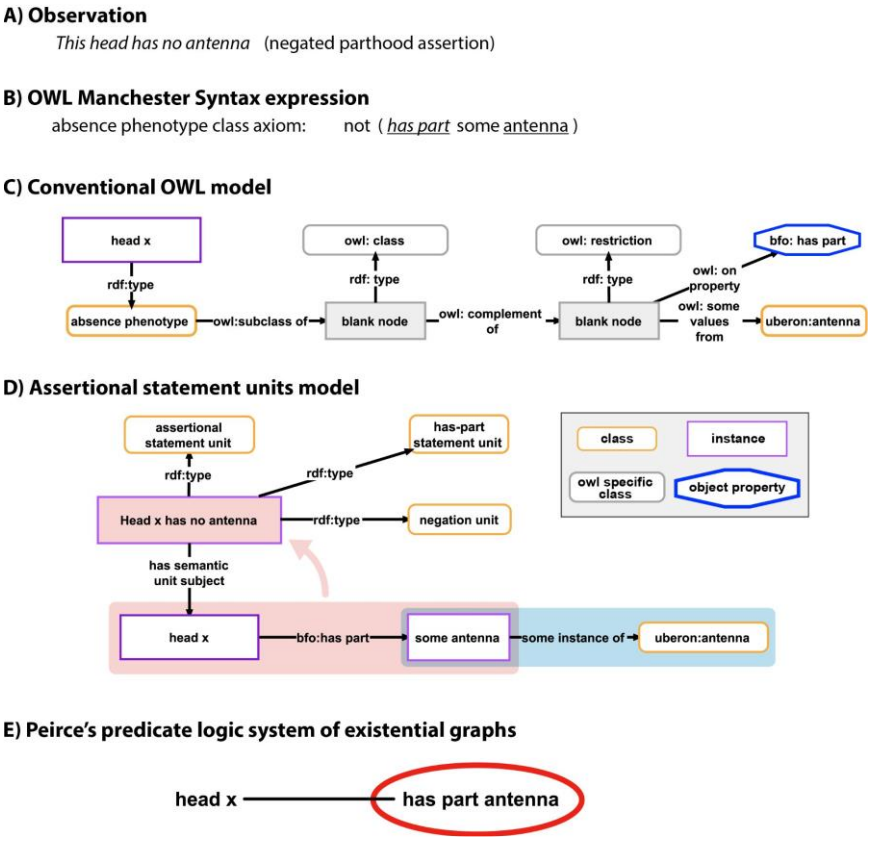


图18：缺失观察、相应断言以及在知识图谱中对它们进行建模的两种替代方式之间的关系。A)关于某个给定头部没有触角这一观测的人类可读陈述。**B)**缺失陈述无法表示为实例之间的关系。因此，A）中的观测必须使用类表达式来表示，而类表达式可以使用曼彻斯特语法来表述。按照这种表示法，头部将是一个类的实例，该类被定义为仅包含自身没有触角作为其部分的实例（“not”和“some”被用作数学表达式）。**C)** 将 A) 和 B) 中的断言转换为映射到 RDF 的 OWL 表达式。请注意，*缺失表型*被定义为一组子类关系和补类限制，涉及两个空白节点。**D)** 同一陈述可以用两个语义单元建模。其中一个对具有部分关系进行建模并将其否定（红色边框框及其动态标签，以及无边框红框中的数据图）。因此，它既是*具有部分关系陈述单元*的实例，也是*断言陈述单元*和*否定单元*的实例。另一个语义单元是*某实例识别单元*的实例，并通过属性*某实例属于*将“某个触角”与触角（UBERON:0000972）关联起来。其数据图显示在蓝框中。二者共同对 A) 中的观测进行建模。这种表示法与 F) 中的表示法相似。**E)** 陈述 “ 这个头部没有触角 ” 以及 A) 中观测的另一种表示法。该表示法使用皮尔士谓词逻辑系统中的存在图。同一性线 —连接短语“头部 *x*”和“具有部分触角”，表示头部 x 有某个触角作为其部分；而环绕后一短语的红圈则通过交叉同一性线来表达否定。为使表示清晰，“头部 *x*”与头部（*UBERON:0000033*）之间的关系未在 *C)* 和 *D)* 中显示。

同样的方法也可用于表示两个实例之间关系的否定。某个给定果实不属于某棵特定橙树这一事实，可以采用与其他否定语句类似的方式进行建模（见图19）。

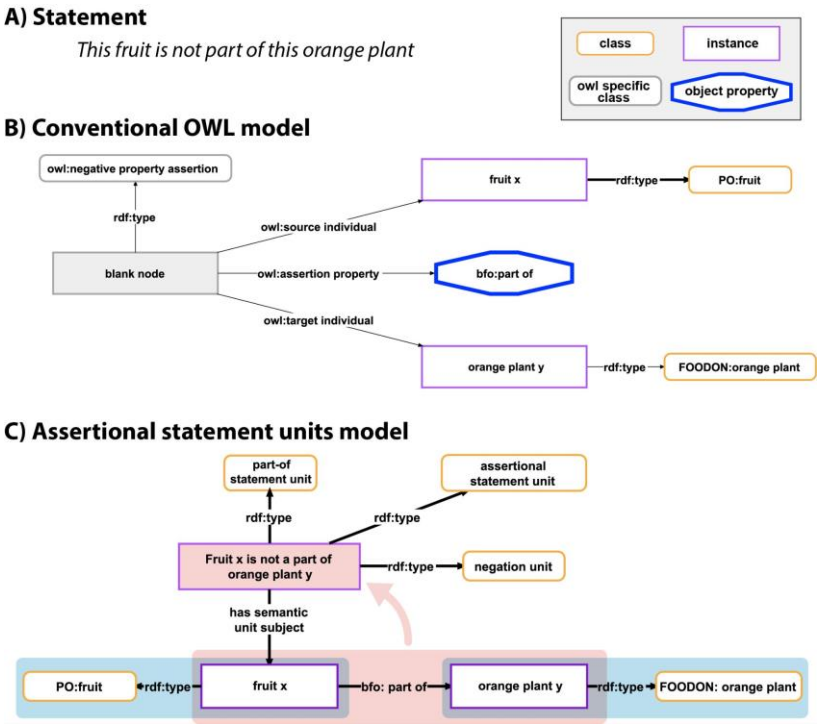


图19：通过应用语义单元对实例之间的否定关系进行建模。A) 一条人类可读语句，说明这个果实不属于这棵橙树。该语句可以通过两种不同方式建模：**B)** 将其表示为映射到 RDF 的 OWL 表达式。请注意，该语句如何被转换为一条否定断言语句，其中包含源、属性和目标规范，将一个实例“果实 x”（PO:0009001）与一个实例“橙树 y”（FOODON:03411339）关联起来，并涉及一个空白节点；或者 **C)** 使用三个语义单元。这两个命名个体识别单元，带有蓝色方框中所示的数据图表明，其中一个宾语是果实，另一个是橙树；而部分关系语句单元（红色边框方框及其动态标签，其数据图显示在无边框的红色方框中）则表明，果实 x 不属于橙树 y，因为它同时实例化了*部分关系语句单元*、*断言陈述单元*和*否定单元*。

从所有这些示例中可以清楚看出，否定（包括缺失陈述）可以通过将相应的语义单元归类为*否定单元*，从而利用语义单元高效地表示。这种通用记法可以与皮尔士的存在关系图（80,81）相比较。在存在关系图中，可以用 同一性线“—”来表达“某物是 A”，记作“—**A**”。同一性线可以理解为表示存在量词（∃*x*）。用一个圈住 A 的圆圈截断这条线，便可表达“某物不是 A”（见图18E）。由此得到的存在关系图具有足够的通用性，能够表示带等号的一阶逻辑（81）。

出于相同的原因，使用 RDF/OWL 表达 **基数限制** 同样颇具挑战性，需要使用一个类表达式，将基数表示为类限制。无需使用相当复杂的类公理来描述例如某个给定的头部恰好有三只眼睛（图 20B），而是可以用两个语义单元来建模相同的信息，并允许扩展某实例识别单元，使其能够通过将其主语与数值 3（例如）关联起来，从而纳入基数限制；关联所用的属性为限定基数（OWL:qualifiedCardinality）（图 20C）。若不使用简单的整数值，而改用一个浮点值范围，并通过单位规范将其限制为计数单位（UO:0000189）和百分比（UO:0000187），甚至还可以指定频率

和范围。具有部分关系陈述单元将“头部 x”与某个语义单元的主语关联起来，该语义单元同时实例化了*某实例识别单元*和*基数限制单元*。

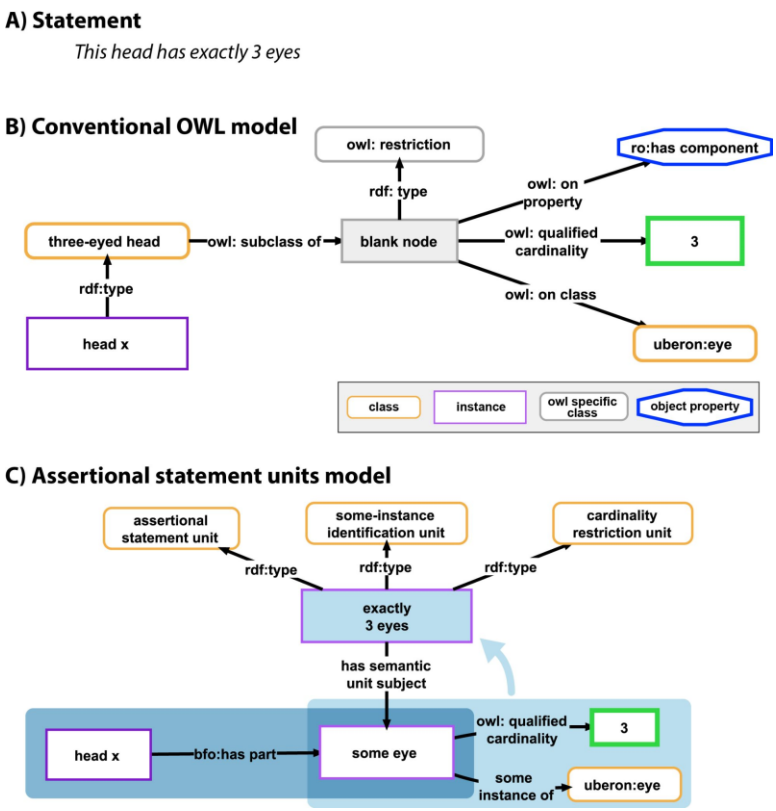


图20：通过应用语义单元对涉及实例的基数限制进行建模。 **A)** 一条可读性强的陈述，说明这个头部恰好有三只眼睛。该陈述可以用两种不同方式建模：**B)** 建模为映射到 RDF 的 OWL 表达式。请注意，“恰好有三只眼睛”如何被转化为某个类的实例，而该类对具有组件属性（RO:0002180）和眼类（UBERON:0000970）施加了基数为 3 的基数限制，因此涉及一个空白节点；或者**C)** 使用两个语义单元，其中一个对具有部分关系进行建模，其数据图显示在深蓝色方框中。另一个语义单元（浅蓝色边框方框及其动态标签，其数据图显示在无边框的浅蓝色方框中）实例化*断言陈述单元*、*某实例识别单元*和*基数限制单元*。

与否定建模的方法类似，语义单元也可用于对知识图谱中的**分歧**进行建模。例如，人物 A 通过命名个体识别单元断言“这个果实是梨果”，而人物 B 不同意这一陈述，那么可以将该分歧建模为一个语义单元，该单元实例化*断言陈述单元*和*分歧单元*，其数据图则指出，人物 A 所断言的命名个体识别单元实例化了*否定单元*（见图 21）。

尽管此处提出的表示法所生成的图比其基于 OWL 的 RDF 等价形式更为简洁，并且由于不存在空白节点，这些图也更易于查询，但仍需开发能够直接使用语义单元的推理器；不过，通过转换为 OWL（见“语义单元的逻辑语义”一节），标准 OWL 推理器将能够对语义单元进行推理，包括否定、基数限制和分歧。

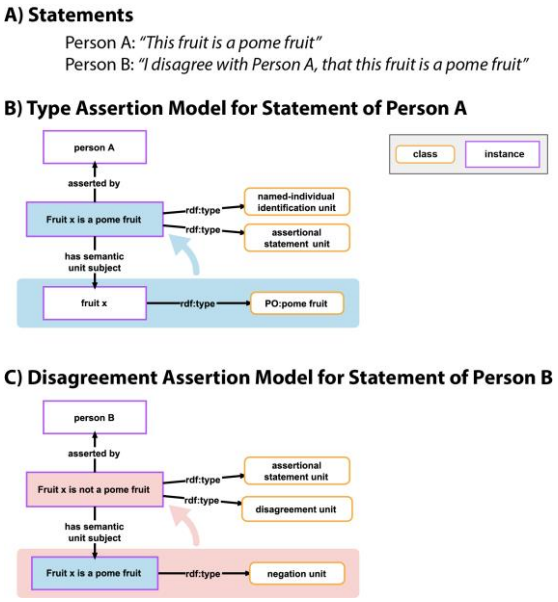


图 21：通过应用语义单元对分歧进行建模。

A）人物 A 声明这是梨果，而人物 B 表示分歧。**B）**断言陈述单元（蓝色边框的框及其动态标签，以及蓝色框中无边框显示的数据图）对人物 A 的陈述进行建模。**C）**分歧单元（红色边框的框及其动态标签，以及红色框中无边框显示的数据图）对人物 B 的陈述进行建模，并实例化*断言陈述单元*和*分歧单元*。其数据图表明，人物 A 的陈述实例化了*否定单元*。

语义单元的逻辑语义

从语义上说，我们将语义单元视为 ABox 与 TBox 语句之间的转换，或者也可以视为逻辑程序与 OWL 公理之间的转换。例如，“Lars 的右手”（FMA:9712）具有部分“Lars 的右拇指”（FMA:24938）是一条 ABox语句。其表征中还包含这样一条信息：该语句表达的是一种“具有部分”类型的公理（具有部分关系陈述单元）。从语义上说，我们将这些陈述单元视为特定形式的本体设计模式的表达；根据陈述单元的类型，该本体设计模式将 ABox 语句转换为 TBox 公理。我们可以使用**关系型本体设计模式**（82）的表达式将这些转换形式化；这类表达式是包含任意实体变量的 OWL 公理。例如，*“具有部分断言陈述单元”*通常可以转换为

?X subclassOf：具有部分 *some ?Y*，

其中，*?X*由语句的主语填充，*?Y*由宾语填充。我们可以将该模式作为字面量添加到陈述单元类中，即将其设为数据类型属性或注释属性。这样就可以使用 OWL 推理器，或者使用 SPARQL 查询，检索所有可依据此设计模式重写的语句；不过，将它们分别指定并采用双重形式体系会更方便。

将复杂公理双重表示为 ABox 语句，并将其转换为 TBox（或 ABox）公理，便可进行**双重类型的推理**。通过转换为 TBox 公理，便能够应用 OWL 推理，因此可以使用任何**OWL 推理器**（或适用于 OWL 配置文件的推理器）

来检验知识图谱的一致性或进行推理。不过，将公理以简化模式表示为 ABox 表示，也支持直接对 ABox 语句进行某种形式的推理，并采用**逻辑编程**的语义；这种推理形式的能力可以超越 OWL 的语义（ 83 ）。尤其是，还可以添加**非单调语句**，以便对偶然陈述或 " 默认规则 " 进行建模，而这些默认规则可能存在例外（84）。例如，语句"通常，手有一个拇指作为其部分"可以通过逻辑程序建模：如果存在一个 Hand（FMA:9712）的实例*x*，且无法证明*x*没有某个 Thumb（FMA:24938）实例作为其部分，那么*x*就有某个 Thumb 实例作为其部分。形式上，这可以表示为如下**答案集程序**：

具有部分*(x,* 拇指*) :- rdf:type(x,* 手*), not lacks-*部分*(x,* 拇指*)*。

然而，这样的陈述具有若干需要解释的不同寻常之处。首先，它不基于 OWL语义，因此最初不会与 OWL 本体中的公理发生交互；具体而言，该陈述单元不会被转换为 OWL 公理，而是转换为一条应用于陈述单元本身的逻辑规则。其次，在答案集语义下，"not"表示弱否定，其预期含义是*lacks-part(x, Thumb)*无法被证明。因此，它编码了一种默认知识，使用这一陈述会导致**非单调推理**：添加一条新陈述，例如*lacks-part(x, Thumb)*，将使先前的推理（*has-part(x, Thumb)*）失效；这与 OWL 的标准语义形成对比，后者是单调的（添加新陈述永远不会使推理失效）。第三，符号*Hand* 和*Thumb* 在此用作个体的符号，而实例属于是实体之间的一种显式关系；而在 OWL 中，*Hand* 和*Thumb* 则是类（对应于一元谓词），实例属于是将个体与类关联起来的内置原语。

为将逻辑程序与 OWL 联系起来，我们假设逻辑程序能够将所有 OWL 实体（类符号、个体符号和关系符号）用作个体。原子公式的形式为 $P(a_1, \ldots, a_n)$，其中 P 是谓词符号，每个 a_i 要么是个体符号，要么是变量符号。不含变量符号的公式称为基元公式。现有多种高效的基化器，例如 Clingo (85)，以及答案集求解器，例如 DLV (86) 或 clasp (87)，它们可以计算完全基元的答案集（或答案集程序的稳定模型）。我们将答案集程序中的谓词符号视为本体设计模式，并将其转换为一组 OWL 公理。以前向方式应用模式是可判定的，而且实现起来很直接：先计算答案集程序完全基元化后的稳定模型，再应用设计模式的正向转换。我们也可以实现从 OWL 到这些模式的逆向转换：对于每个已定义的本体设计模式，通过查询判断该模式是否

可由 OWL 公理蕴涵；借助 OWL 推理器，我们可以在本体 $O^{\vdash}$ 的演绎闭包中查询该模式，并遍历所有 OWL 实体。然而，对于含有 n 个变量、且 OWL 本体中有 |E| 个实体的模式，其复杂度为 O(|E|n)。采用这种方法，语义单元由两部分构成：一条以逻辑程序形式表示的语句，以及一个用于将谓词转换为 OWL 陈述的模式。例如，我们可以将命名个体识别单元（图 8）表示为：

逻辑程序：*rdf:type(Lars'RightHand, fma:hand)*

模式：*rdf:type(Lars'RightHand, fma:手)*

对于命名个体识别单元，两条语句是相同的。更一般地，我们可以使用如下逻辑程序来表述所有命名个体识别单元：

> 命名个体识别单位*(x)*，具有语义主语*(x,y)*，*rdf:type(y, z)*，*rdfs:label(y, l)*，*owl:*命名_个体*(y);*

以及一种将每个谓词转换为相应 OWL 语句的模式。

某实例识别单元在语义上等同于命名个体识别单元，但不会显式引入个体名称。这对应于存在量词的用法，我们可以通过"斯科伦化"消除它；也就是说，我们可以引入一个新的个体名称。这个名称在语义单元的语言中未被显式提及，但存在于语义空间中。

每个实例识别单元需要更复杂的 OWL 转换。在这里，**X 的每一个实例** 指的是 X 的所有实例构成的集合，我们可以借助集合与集体理论 (88)；随后，图 8 中的示例将转换为：

- *rdf:type(everyHand,* 集合*)*
- *owl:SubClassOf(fma:Hand, owl:SomeValuesFrom(member-of, owl:oneOf({everyHand})))*
- *owl:SubClassOf(owl:oneOf(everyHand), owl:AllValuesFrom(*具有成员*, fma:Hand))*

换句话说，*'每只手'* 指的是这样一个集合：其成员包括类 Hand（FMA:9712）的所有实例，且成员仅限于该类的实例。类似地，我们可以翻译基数限制（基数限制单元）；例如，图20中的例子可以用以下（一般）逻辑程序表示：

> 基数限制单位*(x)*，具有语义单位主语*(x,y)*，*owl:qualified-cardinality(y,z)*，*y* 的某个实例是 *w*。

将此模式转换为 OWL，结果为

*rdf:type(cX, owl:intersectionOf(Collection, owl:cardinality(has-member, 3, uberon:eye)))*其中，*cX*是一个新的个体名称（即未在其他地方使用过的名称）。随后，将它与同一示例中关于"头部 X"的"部分关系断言语句单元"结合起来，以断言 ABox语句

part-of(`头部 *X', cX).*

我们可以采用类似的翻译方式，将其他语句形式化。断言陈述单元（图 9）直接翻译为ABox语句，全称陈述单元也采用类似的翻译方式；图 9 中的语句将翻译为 OWL ABox公理

hasPart(everyHand, someThumb)。

基于 OWL语义以及用于形式化 *'每只手'* 和 *'someThumb'* 的公理，可以推导出 OWL 公理

*owl: SubClassOf(fma:Hand, owl:SomeValuesFrom(*具有部分*, Thumb)*

正如预期，图9中的或然陈述单元也被翻译为一条 OWL ABox公理。不过，我们的形式体系还允许表达另一种类型的偶然陈述，即关于原型的陈述（原型偶然语句单元）。这类陈述可用于表述诸如"通常，手具有拇指这一部分"之类的表达，同时允许存在例外（例如因意外事故而失去拇指）。以下陈述表达了这一原型陈述：

具有部分*(x,* 拇指*) :-* 具有部分*(*手*,* 拇指*),* 实例属于*(x,* 手*), not -*具有部分*(x,* 拇指*)*。

采用逻辑编程范式还使我们能够将图 13 所示的**复杂陈述**形式化，这类陈述**无法直接用 OWL 表达**；而通过翻译模式将其翻译为 OWL，则能在与其他 OWL 公理结合时支持 OWL 中的推理。

我们还可以通过使用否定单元来对（经典）**否定**进行建模（图17）。否定会直接添加到语义单元上，我们可以据此定义一个否定断言陈述单元：

否定断言语句单元*(x) :-* 否定单元*(x),* 断言语句单元*(x)*

然后将其转换为 OWL 中的 否定断言语句单元，前提是满足以下逻辑程序：

NegatedAssertionalStatementUnit(x)、*has-semantic-unit-subject(x,y)*、*rdf:type(y,z)*以及 OWL 转换

rdf:type(y, owl:complementOf(z))。

这个例子展示了我们使用逻辑编程的灵活性：否定断言语句单元谓词是根据添加到陈述单元中的两个断言推理得出的，无须显式断言。然而，这种推理方式，尤其是在使用否定单元时，会带来另一项挑战，因为语句可能适用多种模式，而且有时并不容易确定应采用哪种翻译模式。例如，在图17的例子中，如果忽略否定单元，我们就会得出结论：应采用断言陈述单元的模式；而同时应用两种陈述单元的翻译会导致不一致。避免这一问题最简单的方式，是在断言陈述单元的逻辑程序中加入弱否定条件：*not NegatedAssertionalStatementUnit(x)*，或者更简单地写成 *not NegationUnit(x)*；该条件可防止在存在否定单元时触发翻译模式。

此外，逻辑编程范式可用于在语句（视为个体）与其内容（视为命题）之间进行转换。如图 21 所示，语句本身也是独立的单位（断言陈述单元），可使用我们前面介绍的模式将其转换为 OWL 公理。此外，语句还是逻辑程序中的个体，其中可以对语句之间的关系进行建模；这也使得我们能够使用答案集编程来表示论证（89,90）。

语义单元为知识图谱对齐提供了一种新框架

实例化同一语义单元本体类的语义单元包含语义相似的信息。因此，原则上，同一类型的语义单元在其所有实例之间均可进行比较。由此，基于同一组语义单元本体类的不同知识图谱，可以沿着不同表征粒度层级，通过**分步流程**进行对齐和比较。第一步，根据项目组单元的数据图与不同类型主语之间的 **hasSemanticUnitSubject** 关系，以及与不同类型关联语义单元之间的 ***hasAssociatedSemanticUnit*** 关系，对这些数据图进行对齐。下一步，对于每对项目组单元，根据其关联项目单元所属数据图中的主语类型及其关联

陈述单元的类型对数据图进行对齐；而陈述单元又可按类进行对齐，最后还可对齐其单个三元组。这将为改进知识图谱对齐、子图匹配、图比较和图相似度度量方法提供一个新的框架。

管理敏感数据的受限访问

陈述单元及其对应本体类的分类，为识别知识图谱中可能包含敏感数据、因而必须限制访问的子图提供了一个框架。例如，所有与濒危物种出现位置相关的位置数据都不应公开，访问应受到限制，并通过适当的规则进行管理。可以按类识别相应的陈述单元，而是否需要限制对这些单元的访问，则可根据相应物种的威胁等级决定。如果某物种濒临灭绝，对这些类型陈述单元的访问将自动受到限制，而所有其他指涉该物种的语义单元仍可公开访问。同样，任何个人数据都受到隐私政策限制，对相应陈述单元的访问应取决于具体的访问权限。这符合 EOSC 的原则：“尽可能开放，必要时保持封闭”（16）。

结论与未来工作

借助语义单元，我们引入了一种知识图谱中前所未有的资源类型，它显著**增强了知识图谱的整体表达能力**。语义单元将知识图谱组织为可识别且具有语义意义的子图，因此，除了实例、类和属性之外，还代表了一种额外的**表征实体**（91）。**语义单元所表示的资源，其抽象层级高于传统单个三元组中所用资源的低层抽象**。从技术上说，语义单元资源是相应本体类的实例；但从语义上说，它们表示其数据图的内容，因而表示语句或语义和本体上相关的语句集合。因此，借助语义单元，人们可以**从知识图谱中提取具有语义意义的实体**，其抽象程度高于传统三元组中通常使用的资源。

针对引言中讨论的四项挑战，我们可以得出结论：由于每条陈述都组织在各自的陈述单元中，而每个陈述单元又实例化相应的陈述单元本体类，因此可以为每个此类类指定一个**数据模式**（例如，以形状的形式）以及相应的**CRUD 查询模式**。通过将数据模式链接到相应的陈述单元类，每个陈述单元的 UPRI 及其归属的相应

陈述单元类也会**指称底层数据模式，从而通过确保模式互操作性，提升其数据对 FAIR 原则的遵循程度**。由于陈述单元对知识图谱进行划分，使每个三元组都恰好属于一个陈述单元，因此知识图谱中的所有数据都会包含其底层数据模式的信息。指向底层数据模式的指称构成了有价值的元数据，可确保同一类的所有陈述单元实例之间的互操作性，因为它们都基于相同的数据模式（挑战 *1*）。

CRUD 查询模式集不仅能为领域专家提供读取查询，以便在整个数据图中搜索相应类的实例，还能为开发者提供创建、更新和删除查询。开发者可以使用这些查询，因此不再需要学习图查询语言（挑战*2*）。

通过将整个数据图组织成不同的语义单元，从而形成表征粒度各异且语义明确的子图，语义单元还解决了在知识图谱中难以作出**关于陈述的陈述**这一问题（挑战*3*）。语义单元为组织图提供了一种高效方式，使用户能够直观地作出关于陈述的陈述，同时也提供了清晰、直接的实现方案，支持开发相应的查询和相关工具。此外，与 RDF-star 不同，语义单元还允许将同一陈述的不同实例组织为不同的语义单元，并使它们实例化同一个语义单元类，从而加以区分。这些实例的主语和宾语相同，但各自的语义单元资源不同。

最后，通过将断言式、偶然式（包括原型式）和全称陈述单元区分为陈述单元的顶层类别，引入某个实例资源和每个实例资源，并引入一种不涉及空白节点且在映射到 RDF 时比等效 OWL 表示法更简单的表示法，语义单元提供了一个概念框架，借助它可以弥合 RDF/OWL 与属性图之间的概念鸿沟，为偶然陈述和原型陈述提供形式语义，并使全称陈述以及特定类公理也能成为知识图谱论域的一部分（挑战*4*）。换言之，**语义单元显著提升了知识图谱的整体表达能力**。

未来，我们希望扩展语义单元这一概念，为每个语义单元类纳入相应的数据模式和查询模式。结合其他功能，这就形成了我们所称的 **知识图谱构建块**（KGBB）（另请参见知识图谱单元（79），其中首次

讨论了类似的理念）。KGBB 是一个与语义单元本体类关联的小型模块。每个语义单元类都需要拥有其自身关联的 KGBB。KGBB 提供管理属于相应类型语义单元的数据所需的全部信息。每个 KGBB 都为存储其关联类型语义单元的数据提供图模型，并可据此推导出 CRUD 查询模式，从而产生 FAIR 数据和元数据。KGBB 通过以各种格式提供数据访问，**将数据存储与数据访问解耦**；同时，它们通过提供不同的显示模板，**将数据显示与数据存储解耦**。最终，KGBB 的理念是：每个 KGBB 都能独立于其他 KGBB 运行；将其公开放入代码库中，以便他人复用；领域专家能够以尽可能少的工作组合多个 KGBB，并定义它们之间可能的交互，从而搭建自己的知识图谱应用，而无需成为程序员或掌握语义学知识。**KGBB 编辑器**将使不具备语义学、编程或图查询语言背景的领域专家能够描述新的 KGBB。**KGBB 应用引擎**将利用知识图谱应用中各个 KGBB 提供的信息，并通过相应的 API 与持久化层和表示层通信，从而**不仅将人类可读的数据展示与数据存储解耦，还将数据访问与数据存储解耦，并将数据存储与存储技术解耦**。

其中一位作者已开发出首个最小可行产品，展示了如何利用 KGBB 和语义单元来管理知识图谱（92）。该产品基于 Neo4j，通过网页界面和用户输入获取内容。它是一个小型知识图谱应用，用于记录学术出版物中的断言，并以示例方式让用户描述其中的一些内容，但不侧重于描述出版物的书目元数据）。每篇论文都作为独立的项目组单元表示，其中涉及的断言由陈述单元涵盖，并与项目单元和粒度树单元相关联。该示例基于 Python 和 Flask/Jinja2 构建，并通过https://github.com/LarsVogt/Knowledge-Graph-Building-Blocks公开提供。

我们认为，将语义单元与知识图谱构建模块相结合，将为跨众多利益相关方共享数据提供一个框架，帮助来自不同领域的数据提供者彼此互联，同时由领域专家或机构掌管数据管理，从而确保其技术自主权（遵循 Barend Mons 提出的数据访问，而非数据共享 (7)）。我们还认为，得益于其模块化特性，该框架能够提高软件开发者和缺乏语义学专业知识的领域专家访问知识图谱的便利性，进而推动数据和元数据的 FAIR 化。这正是让所有数据实现 FAIR 所迫切需要的；我们相信，这将提高我们应对气候变化和生物多样性丧失的胜算。

致谢

我们感谢 Werner Ceusters、Nico Matentzoglu、Manuel Prinz、Marcel Konrad、Philip Strömert、Roman Baum、Björn Quast、Peter Grobe、István Míko、Manfred Jeusfeld、Manolis Koubarakis、Javad Chamanara 和 Kheir Eddine 与我们讨论了文中提出的一些观念。Lars Vogt 获得了 ERC H2020 项目“ScienceGraph”（819536）的资助。本文中的所有论证和陈述均由我们负责。

作者贡献

LV：提出了语义单元的概念并撰写了稿件。TC：参与撰写了关于纳米出版物的章节。RH：贡献了形式语义。所有作者：审阅了稿件。

参考文献

1. Adam, K., Hammad, I., Adam, M., 等（2015）《大数据分析与存储》。*2015*年运营卓越与服务工程国际会议论文集，奥兰多，佛罗里达州，美国，*2015*年*9*月*10*日至*11*日，第648–659页。2. Marr, B.（2018）我们每天会创造多少数据？人人都该了解的惊人统计数据。我们每天会创造多少数据？人人都该了解的惊人统计数据；（2018）。3. 数据永不眠 5.0。4. Idrees, S.M.、Alam, M.A. 和 Agarwal, P.（2018）大数据及其挑战研究。国际信息技术杂志，**11**，841–846。5. 联合国大会（2015）改变我们的世界：2030年可持续发展议程。改变我们的世界：*2030*年可持续发展议程；（2015）；第1–35页。6. 国际科学理事会（ISC）（2019）科学作为全球公共产品：ISC行动计划，2019—2021年。科学作为全球公共产品：*ISC*行动计划，*2019—2021*年；国际科学理事会，（2019）。7. Mons, B.（2019）CODATA主席Barend Mons致辞。*CODATA*主席*Barend Mons*致辞；（2019）。8. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., 等（2016）科学数据管理与管理责任的FAIR指导原则。《科学数据》，**3**，160018。9. P. Bryan Heidorn（2008）揭开科学长尾中的暗数据。《图书馆趋势》，**57**，280–299。10. Baker, M.（2016）1500名科学家揭示可重复性现状。《自然》，**533**，452–454。11. Lin, D., Crabtree, J., Dillo, I., 等（2020）数字存储库的TRUST原则。《科学数据》，**7**，144。

12. The FAIR Guiding Principles for scientific data management and stewardship | Scientific Data 互联网。检索于 Mayris, P., Bakker, J.-Y., Bruce, R., 等（2016）实现欧洲开放科学云。实现欧洲开放科学云; (2016) ; 第24页。14. Hasnain, A. 和 Rebholz-Schuhmann, D.（2018）根据五星开放数据原则评估FAIR数据原则。载于 Gangemi, A., Gentile, A. L., Nuzzolese, A. G., 等（编），语义网：*ESWC 2018*卫星活动。*ESWC 2018*。《计算机科学讲义》，第*11155*卷。，施普林格，第469–477页。15. 欧盟委员会研究与创新总司和EOSC执行委员会（2021）EOSC互操作性框架：EOSC执行委员会FAIR与架构工作组报告。*EOSC*互操作性框架：*EOSC*执行委员会*FAIR*与架构工作组报告。; 出版物办公室，卢森堡，（2021）。16. 欧盟委员会FAIR数据专家组（2018）让FAIR成为现实。让*FAIR*成为现实; (2018)。17. Vogt, L., Baum, R., Bhatty, P., 等（2019）SOCCOMAS：一种运用知识图谱并基于语义编程的FAIR网页内容管理系统。数据库，**2019**，1–22。18. Blumauer, A.（2018）知识图谱——在日益复杂的世界中串联点滴。知识图谱——在日益复杂的世界中串联点滴; (2018)。19. Singhal, A.（2012）知识图谱简介：关注事物，而非字符串。知识图谱简介：关注事物，而非字符串**（2012）**。20. Hogan, A., Blomqvist, E., Cochez, M., 等（2021）知识图谱。《*ACM*计算调查》，**54**，1–37。21. Bonatti, P.A., Decker, S., Polleres, A., 等（2019）知识图谱：语义网上知识表示的新方向。《达格施图尔研讨会报告》，**8**，29–111。22. Abiteboul, S.（1997）半结构化数据查询。载于 Afrati, F., Kolaitis, P.（编），数据库理论 — *ICDT'97*，Springer Berlin Heidelberg，德国柏林、海德堡，第1–18页。23. Angles, R. 和 Gutierrez, C.（2008）图数据库模型综述。《*ACM*计算调查》，**40**，1–39。24. Angles, R., Arenas, M., Barceló, P., 等（2017）图数据库现代查询语言基础。《*ACM*计算调查》，**50**，1–40。25. Hitzler, P., Krötzsch, M., Parsia, B., 等（2012）OWL 2 Web本体语言入门（第二版），W3C 2012年12月11日建议标准。OWL 2 Web本体语言入门（第二版），W3C 2012年12月11日建议标准。**（2012）**。26. Stutz, P., Strebel, D. 和 Bernstein, A.（2016）Signal/Collect12：数秒内处理大型图。语义网，**7**，139–166。27. Wang, Q., Mao, Z., Wang, B., 等（2017）知识图谱嵌入：方法与应用综述。《*IEEE*知识与数据工程汇刊》，**29**，2724–2743。28. Ceusters, W.（2020）指称对象追踪在生物医学信息学中的地位。术语学、本体及其实现，施普林格自然，第50页。29. Ceusters, W., Elkin, P. 和 Smith, B.（2007）电子健康记录和生物医学本体中的阴性发现：一种实在论方法。《国际医学信息学杂志》，**76**，S326–S333。30. Vogt, L.（2021）电子科学时代的FAIR数据表示：以表型描述为例，对经验数据的基于实例和基于类的语义表示进行比较。《生物医学语义学杂志》，**12**，20。

31. Bandrowski, A., Brinkman, R., Brochhausen, M., 等（2016）生物医学调查本体。*PLoS ONE*，**11**，1–19。32. Knublauch, H. 和 Kontokostas, D.（2017）形状约束语言（SHACL）——W3C 2017年7月20日建议标准。形状约束语言（*SHACL*）——*W3C 2017*年*7*月*2*日建议标准; (2017)。33. Booth, D. 和 Wallace, E.（2019）专题X：EasyRDF。专题X：EasyRDF。第二届美国语义技术研讨会 *2019*（**2019**）。34. Hartig, O.（2017）RDF*和SPARQL*基础。*CEUR*研讨会论文集*1912*。35. Arndt, D., Broekstra, J., DuCharme, B., 等。RDF-star和SPARQL-star草案社区组报告，2021年7月1日。*RDF-star*和*SPARQL-star*草案社区组报告，*2021*年*7*月*1*日。36. Frey, J., Müller, K., Hellmann, S., 等（2019）RDF存储库中的元数据表示评估。《语义网杂志》，**10**，205–229。37. Mungall, C.J.（2019）RDF*和属性图语义的拟议策略。RDF*和属性图语义的拟议策略（**2019**）。38. Mungall, C.J.（2017）Shadowboxing：在ABox中映射TBox。Shadowboxing：在ABox中映射TBox（**2017**）。39. Hogan, A., Arenas, M., Mallea, A., 等（2014）关于空白节点，你一直想知道的一切。《网络语义学杂志》，**27–28**，42–69。40. Neumann, T. 和 Moerkotte, G.（2011）特征集：对包含多个连接的RDF查询进行精确基数估计。*2011*年*IEEE*第*27*届国际数据工程会议，IEEE，德国汉诺威，第984–994页。41. Papastefanatos, G., Meimaris, M. 和 Vassiliadis, P.（2021）基于RDF的知识图谱的关系模式优化。《信息系统》，101754。42. Collarana, D., Galkin, M., Traverso-Ribón, I., 等（2017）MINTE：以语义方式整合RDF 图。第七届国际网络智能、挖掘与语义会议论文集 - *WIMS '17*，ACM出版社，美国纽约州纽约市，第1–11页。43. Endris, K.M., Galkin, M., Lytra, I., 等（2018）通过衔接RDF分子模板查询互链接数据。通过衔接RDF分子模板查询互链接数据（**2018**）。44. Vogt, L.（2019）组织表型数据——一种用于解剖结构的语义数据模型。《生物医学语义学杂志》，**10**，12。45. Smith, B., Almeida, M., Bona, J., 等（2015）基础形式本体 2.0。46. Schulz, S. 和 Jansen, L.（2013）生物医学知识表示中的形式本体。《*IMIA*医学信息学年鉴 *2013*》，**8**，132–46。47. Schulz, S., Stenzhorn, H., Boekers, M., 等（2009）形式本体的优势与局限

s

在生物医学领域。健康传播、信息与创新电子期刊，**3**，31–45。48. Rector, A.、Stevens, R. 和 Drummond, N.（2009）是什么导致了肺炎？论 OWL 中"may"的标准语义。*CEUR*研讨会论文集，**432**。49. Dodds, L. 和 Davis, I.（2011）关联数据模式：用于建模、发布和使用关联数据的模式目录。关联数据模式：用于建模、发布和使用关联数据的模式目录

g,

发布和使用关联数据；（2011 年）。50. Stevens, R. and Sattler, U.（2014）用 OWL 中的树来引导你的思路：为什么有些事情你想说却不能说。用 OWL 中的树来引导你的思路：为什么有些事情你想说却不能说。本体论与知识博客（**2014 年**）。

51. W3C: OWL 2 Web Ontology Language Manchester Syntax (Second Edition). 52. Balhoff, J.P., Mikó, I., Yoder, M.J. 等（2013）应用于新喀里多尼亚旗腹蜂（膜翅目：旗腹蜂科）的物种描述语义模型。*系统生物学*，**62**，639–659。53. Bittner, T. 和 Smith, B.（2001）颗粒划分分类学。载于 Montello, D. R.（编），*空间信息理论：地理信息科学基础*，《计算机科学讲义》第*2205*卷，施普林格，柏林，第16页。54. Smith, B. 和 Bittner, T.（2001）粒度、模糊性与近似的统一理论。*COSIT*空间模糊性、不确定性与粒度研讨会论文集。，第102卷，第39页。55. Bittner, T. 和 Smith, B.（2003）颗粒划分理论。载于 Duckham, M.、Goodchild, M. F.、Worboys, M. F.（编），*地理信息科学基础*，泰勒-弗朗西斯图书公司

伦敦，第117–149页。56. Keet, C.M. (2008)《粒度形式理论——通过粒度增强生物学与应用生命科学信息系统》。博士论文，博尔扎诺自由大学，博尔扎诺。57. Vogt, L. (2010) 生物材料实体的时空结构粒度。《*BMC*生物信息学》，**11**。58. Vogt, L. (2019) 层级与构成模块——迈向生命科学领域粒度框架。《生物医学语义学杂志》，**10**，1–29。59. Gayo, J.E.L., Prud'hommeaux, E., Staworko, S., 等 (2015) 迈向一种基于正则表达式导数的RDF验证语言。*CEUR*研讨会论文集，**1330**，197–204。60. Staworko, S., Boneva, I., Gayo, J.E.L., 等。ShEx用于RDF的复杂性与表达能力。18。61. Lupp, D.P., Hodkiewicz, M. 和 Skjæveland, M.G. (2020) 工业资产维护模板库：实现可扩展且易维护本体的方法论。第*13*届可扩展语义网知识库系统国际会议（*SSWS 2020*），与国际语义网会议（*ISWC 2020*）同期举行，第49–64页。62. Skjæveland, M.G., Lupp, D.P., Karlsen, L.H., 等 (2018) 借助合理的本体模板实现实用的本体模式实例化、发现与维护。载于 Vrandečić, D., Bontcheva, K., Suárez-Figueroa, M. C., 等（编），*语义网– ISWC 2018*，《计算机科学讲义》，Springer International Publishing，查姆，第11136卷，第477–494页。63. Mons, B. 和 Velterop, J. (2009) 电子科学时代的纳米出版物 电子科学时代的纳米出版物。科学论述中的语义网应用研讨会（*SWASD 2009*）。64. Groth, P., Gibson, A. 和 Velterop, J. (2010) 纳米出版物剖析。信息服务与使用，**30**，51–56。65. Kuhn, T., Taelman, R., Emonet, V., 等 (2021) 通过去中心化纳米出版物服务实现语义微贡献。*PeerJ*计算机科学，**7**，e387。66. Giachelle, F., Dosso, D. 和 Silvello, G. (2021) 在万维网上搜索、访问和探索生命科学纳米出版物。*PeerJ*计算机科学，**7**，e335。67. Kuhn, T., Willighagen, E., Evelo, C., 等 (2017) 对不断变化的关联数据进行可靠的粒度化引用。《计算机科学讲义》（包括《人工智能讲义》和《生物信息学讲义》子系列），第10587卷 LNCS，第436–451页。68. Clark, T., Ciccarese, P.N. 和 Goble, C.A. (2014) 微型出版物：生物医学交流中主张、证据、论证与注释的语义模型。《生物医学语义学杂志》，**5**，28。

69. Pinker, S. (1997)《心智如何运作》。《心智如何运作》；W.W. Norton & Company, Inc.，（1997）。70. Herre, H. (2010) 部分整体系统本体论：一种逻辑方法。本体论的理论与应用：哲学视角，Springer Netherlands，多德雷赫特，第57–82页。71. Azzam, A., Polleres, A., Fernández, J.D., 等 (2022) smart-KG：用于查询知识图谱的基于分区的关联数据片段。语义网——互操作性、可用性与应用，**审稿中**。72. Vogt, L. (2011) 符号与术语：科学在语言与感知之间的处境。*Bionomina*，**4**，1–41。73. Smith, B., Kusnierczyk, W., Schober, D., 等 (2006) 迈向生物医学领域本体研究与开发的参考术语。载于 Bodenreider, O.（编），*KR-MED 2006*论文集，《健康技术与信息学研究》第*124*卷，IOS Press，第57–66页。74. Vogt, L., Mikó, I. 和 Bartolomaeus, T. (2022) 解剖结构与生物学中的类型概念表明，本体必须适应研究的诊断需求。《生物医学语义学杂志》，**13**，27

.

75. Marconi, D.（1995）《词汇能力结构》。《亚里士多德学会会刊》，**95**，131–150。76. Smith, B.（2004）《生物分类逻辑与生物医学本体的基础》。《空间认知与计算》，**25**，25–29。77. Stamos, D.N.（2005）《达尔文之前的分类学与本质主义——答 Mary Winsor》。《生物学与哲学》，**20**，79–96。78. Johansson, I.（2006）《生物信息学与生物现实》。《生物医学信息学杂志》，**39**，274–287。79. Vogt, L., D'Souza, J., Stocker, M., 等（2020）《利用知识图谱单元在学术知识图谱中表示研究贡献》。*2020*年*ACM/IEEE*数字图书馆联合会议论文集，ACM，中国线上活动，第107–116页。80. Peirce, C.S. 存在图 - CP 4.372-417。存在图 - CP 4.372-417。81. Sowa, J.F.（2011）《皮尔士关于存在图的教程》。《符号学》，**2011**。82. Hoehndorf, R., Oellrich, A., Dumontier, M., 等（2010）《关系即模式：弥合 OBO 与 OWL 之间的鸿沟》。《*BMC*生物信息学》，**11**，441。83. Lifschitz, V.（2019）《答案集编程》。答案集编程；Springer International Publishing，查姆，（2019）。84. Hoehndorf, R., Loebe, F., Kelso, J., 等（2007）《在生物医学本体中表示默认知识：解剖结构与表型本体集成的应用》。《*BMC*生物信息学》，**8**，377。85. Gebser, M., Kaminski, R., Kaufmann, B., 等（2019）《使用 clingo 进行多轮次 ASP 求解》。*TheoryPract. Log. Program.*，**19**，27–82。86. Adrian, W.T., Alviano, M., Calimeri, F., 等（2018）《ASP 系统 DLV：进展与应用》。《人工智能》，**32**，177–179。87. Gebser, M., Kaufmann, B., Neumann, A., 等（2007）《clasp：一种冲突驱动答案集求解器》。逻辑编程与非单调推理，Springer Berlin Heidelberg，柏林，海德堡，第260–265页。88. Wood, Z. 和 Galton, A.（2009）《集体现象分类学》。《应用本体学》，**4**，267–292。89. Egly, U., Alice Gaggl, S. 和 Woltran, S.（2010）《论证框架的答案集编程编码》。

《论证与计算》，**1**，147–177。90. Toni, F. 和 Sergot, M.（2011）《论证与答案集编程》。第164–180页。91. Rector, A., Schulz, S., Rodrigues, J.M., 等（2019）《超越 Gruber：今日的“本体”》

s

生物医学信息系统与 OWL 的局限性。《生物医学信息学杂志》*X*，**2**，1–15。92. Vogt, L.（2022）《采用语义单元的 FAIR 知识图谱——原型》。采用语义单元的 FAIR 知识图谱——原型（**2022**）。