

# Knowledge Graph Representation of Zeolitic Crystalline Materials

Aleksandar Kondinski<sup>1</sup>, Pavlo Rutkevych<sup>1</sup>, Laura Pascazio<sup>1</sup>, Dan N. Tran<sup>1</sup>,  
Feroz Farazi<sup>2</sup>, Srishti Ganguly<sup>1</sup>, Markus Kraft<sup>1,2,3,4</sup>

released: April 26, 2024

<sup>1</sup> CARES  
Cambridge Centre for Advanced  
Research and Education in Singapore  
1 Create Way  
CREATE Tower, #05-05  
Singapore, 138602

<sup>2</sup> Department of Chemical Engineering  
and Biotechnology  
University of Cambridge  
Philippa Fawcett Drive  
Cambridge, CB3 0AS  
United Kingdom

<sup>3</sup> School of Chemical  
and Biomedical Engineering  
Nanyang Technological University  
62 Nanyang Drive  
Singapore, 637459

<sup>4</sup> The Alan Turing Institute  
2QR, John Dodson House  
96 Euston Road  
London, NW1 2DB  
United Kingdom

Preprint No. 321



---

*Keywords:* zeolites, crystal information, chemical species, Knowledge graphs, question-answering system

**Edited by**

Computational Modelling Group  
Department of Chemical Engineering and Biotechnology  
University of Cambridge  
Philippa Fawcett Drive  
Cambridge, CB3 0AS  
United Kingdom

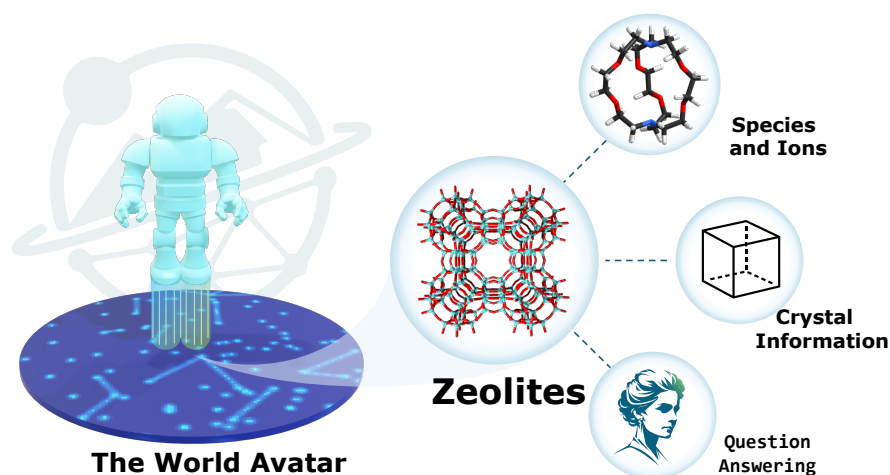
**E-Mail:** [mk306@cam.ac.uk](mailto:mk306@cam.ac.uk)

**World Wide Web:** <https://como.ceb.cam.ac.uk/>



## Abstract

Zeolites are complex and porous crystalline inorganic materials that serve as hosts for a variety of molecular, ionic and cluster species. Formal, machine-actionable representation of this chemistry presents a challenge as a variety of concepts need to be semantically interlinked. This work demonstrates the potential of knowledge engineering in overcoming this challenge. We develop ontologies OntoCrystal and OntoZeolite, enabling the representation and instantiation of crystalline zeolite information into a dynamic, interoperable knowledge graph called The World Avatar (TWA). In TWA, crystalline zeolite instances are semantically interconnected with chemical species that act as guests in these materials. Information can be obtained via custom or templated SPARQL queries administered through a user-friendly web interface. Unstructured exploration is facilitated through natural language processing using the Marie System, showcasing promise for the blended large language model – knowledge graph approach in providing accurate responses on zeolite chemistry in natural language.



## Highlights

- Ontological knowledge representation of crystal and zeolite information.
- Instantiation and interlinking of zeolite, crystal and non-framework species.
- Cross-domain queries and comparisons enhanced via a web interface.
- High-accuracy question answering facilitated by large language models.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Background</b>	<b>5</b>
2.1	Zeolite Architectures and their Chemistry . . . . .	5
2.2	Crystallographic Information . . . . .	7
2.3	Digital Chemistry in The World Avatar . . . . .	8
<b>3</b>	<b>Methodology</b>	<b>9</b>
3.1	OntoZeolite . . . . .	9
3.2	OntoCrystal . . . . .	10
3.3	OntoSpecies . . . . .	11
3.4	Data Curation . . . . .	12
3.5	TWA Integrated Query Interface . . . . .	13
<b>4</b>	<b>Results and Discussion</b>	<b>13</b>
4.1	Overview on TWA Zeolitic Instances . . . . .	13
4.2	Custom SPARQL-based Requests . . . . .	15
4.3	Web-assisted Data Exploration . . . . .	16
4.3.1	Queries for Cross-Framework Comparisons . . . . .	16
4.3.2	Structured Query of Framework Properties . . . . .	17
4.3.3	Structured Identification of Reference XRD Powder Patterns . . . . .	17
4.3.4	Question-Answering based using the Marie System . . . . .	20
<b>5</b>	<b>Summary and conclusions</b>	<b>21</b>
	<b>References</b>	<b>23</b>
	<b>Supporting Information (SI)</b>	<b>31</b>
SI.1	Ontology Design and Data Details . . . . .	31
SI.2	NLP Workflow, Questions and Responses . . . . .	34
SI.3	Structured Queries and Responses . . . . .	52
SI.4	Description Logic . . . . .	55

# 1 Introduction

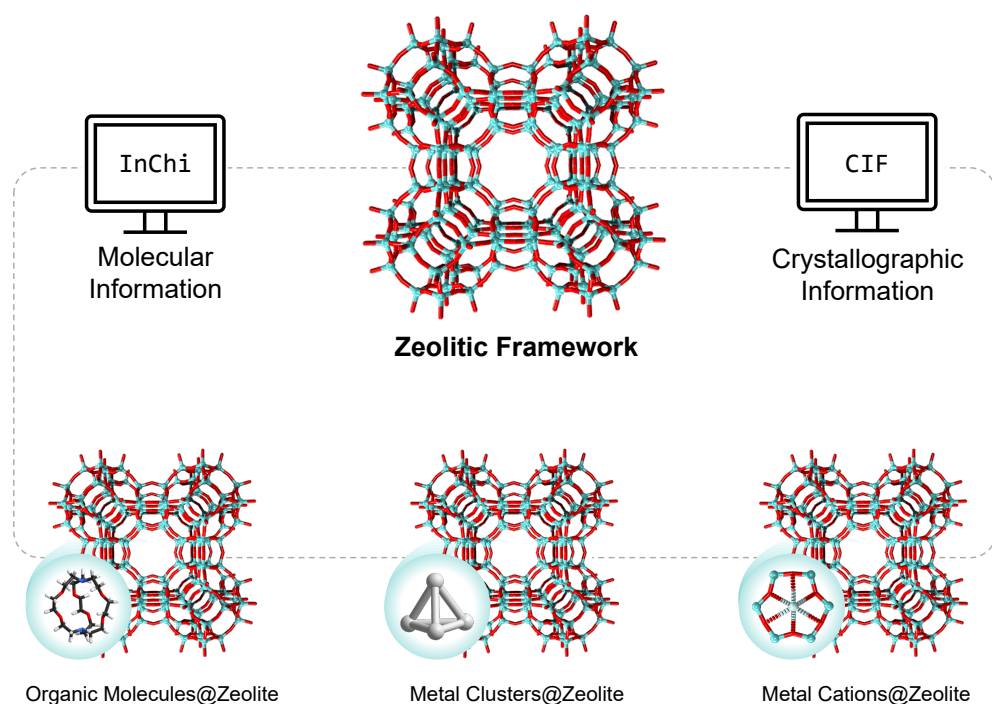
Zeolites are mesoporous inorganic materials which have been of scientific interest since their first description by Fredrik Cronstedt in 1756 [29, 30, 53]. Ancient applications of naturally occurring, mineralogical zeolites include water purification and use as construction materials [30, 73]. Owing to their porosity, fine-tuning of chemical composition, size and topology of the internal channels and cavities, zeolites have been highly relevant in catalysis [20, 81] and separation technologies [62, 84]. Much of the interest in zeolites has been driven by their applications in domains such as crude oil cracking based on shape-selective Brønsted-acid catalysis [63], separations of hydrocarbons [80] and the removal of water/CO<sub>2</sub> from natural gas [46]. Besides these energy-related domains, zeolites find further applications in ion exchange [36, 48] and O<sub>2</sub>/N<sub>2</sub> gas operations from air [37], while new directions include the development of batteries and upcycling of carbon dioxide technologies [89].

The porosity aspect of zeolites was inferred when, upon heating, certain mineralogical aluminium silicates released water vapour [29, 30]. In addition to water molecules, zeolites are recorded to store a variety of other chemical species, including clusters and counter ions. Plenary zeolitic frameworks are typically described as having an ideal generic empirical formula [TO<sub>2</sub>]<sub>n</sub>, where the T-atom is a tetrahedrally coordinated, framework-building element. Aluminosilicates are an example where the positions of the T-atoms are shared between T and T' atoms, while the overall framework zeolite exhibits general formula [T'<sub>x</sub>T<sub>1-x</sub>O<sub>2</sub>]<sub>n</sub>. To completely balance the charge of the two oxo ligands per empirical formula unit, the T/T' atoms are expected to be four valent (*e.g.* Si<sup>4+</sup> or Ge<sup>4+</sup>). However, when framework building centres with other oxidation states participate (*e.g.* Al<sup>3+</sup> or P<sup>4+</sup>S), the overall formal charge of the framework building element components may not be neutral, and thus it may need to be balanced by countercations which find a way in the structure through the network of channels and cavities. In this regard, most of the zeolite framework building elements are *p*-blocks (*e.g.*, Al, Si, Ga, Ge, P, Sn), *s*-block (*e.g.*, Li, Be), or *d*-block (*e.g.*, Ti, Fe). Oxygen atoms are the predominant complementary element for building zeolitic frameworks; however, other atoms such as N, S, or Se may take the position of oxygen in the construction of zeolitic materials [86].

Zeolites precedent the development of other porous reticular materials which obtain a broad prominence nowadays [82]. However, they still retain an enormous interest and fascination owing to their stability, market availability and industrial applications. Computational approaches have been expanding the frontier of research, especially in solving problems for which experimental design and validation can be challenging [71, 78]. With the emergence and accessibility of applied AI, the field has been further advanced simply through data intelligence [19, 31, 52, 68]. Similarly to medical and drug development research, zeolite chemistry is highly interconnected to domains that may not be considered purely chemical in nature [34]. Modelling of the interconnected nature is important to fully capitalise on machine intelligence and advance the field. In this regard, zeolite chemistry combines abstract aspects such as tiling of space and generic framework topologies [47], with crystallographic information, and species/counterion information with its own chemistry in pores and framework directing effects [17, 69].

Over the past decade, our group has investigated the intersection of knowledge engineer-

ing (also known as knowledge AI) and chemistry [42]. Knowledge engineering often relies on semantic web technology that enables efficient machine actionable retrieval and navigation of interconnected information, coupled with dynamic knowledge growth and decision-making facilitated by agent reasoning [9, 74]. In terms of chemical and materials informatics, zeolite chemistry overarches chemical and crystalline material concepts, typically described in different data formats (see Figure 1), making it a subject of fundamental and practical interest. Further on, zeolites are involved in forms of “host-guest” chemistry, and thus their semantic representation is an effort towards developing more general models for simultaneous multi-component information representation in digital chemistry [43].



**Figure 1:** *In terms of information modelling, zeolite chemistry bridges information related to framework topologies, chemical species and crystalline materials.*

In this study, we address the challenge of making zeolite chemistry machine-actionable and subsequently ensure that information can be retrieved in a structured and unstructured manner. This implies that information on zeolite material instances is integrated with information on zeolite topologies and their construction, crystalline information and information on non-framework chemical species functioning as guest or charge-balancing ions inside the framework cavities. These types of information are currently found through different research data resources (see Section 3.4 for more details), and face interoperability challenges. To overcome these challenges, in this work, we apply knowledge engineering to develop two interconnected ontologies, namely “OntoZeolite” and “OntoCrystal”, that deal with zeolitic and crystalline information, respectively. Concepts of these ontologies are semantically interconnected with “OntoSpecies” ontology, [60] that has been previously developed by us and used in the semantic representation of chemical species relevant

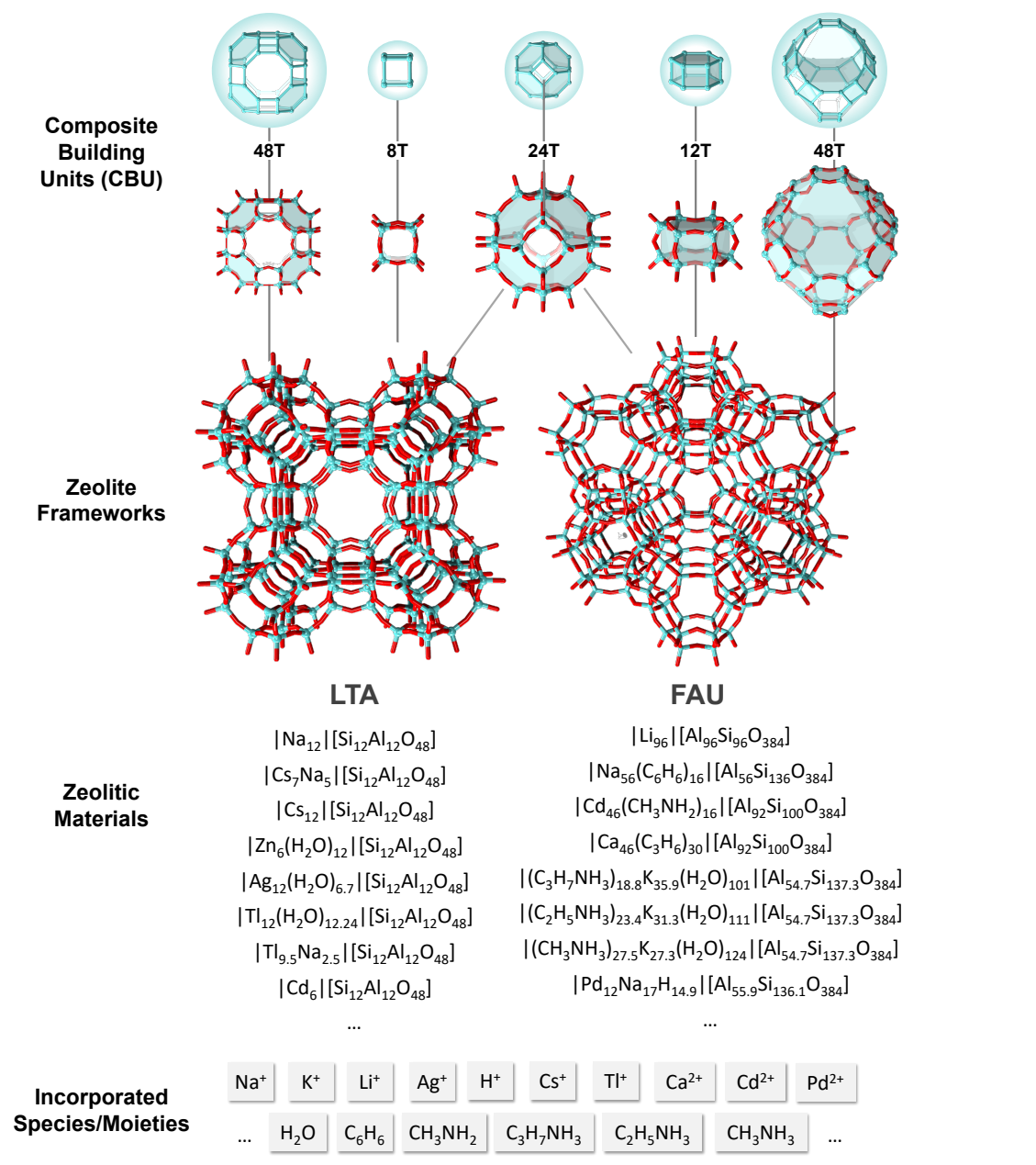
in domains such as chemical kinetics [26, 27], reticular chemistry [41], and experiment automation [7, 8]. Following the integration of the new ontologies with the overall semantic world model of The World Avatar (TWA), we instantiate and interconnect curated zeolite, crystal and species data. Using tailored SPARQL queries, we showcase how interconnected information that is necessary for answering complex chemistry questions can be seamlessly retrieved. Using the TWA capability for question-answering (QA) through its “Marie” system, we open the possibility of zeolite information query using natural language. The application of large language models (LLMs) in chemistry has attracted attention for their potential utility, yet the persistent challenge remains in accurately assessing their performance [18, 22, 38]. Therefore, using Marie herein we provide a blended approach combining the accuracy of knowledge graphs with the natural language understanding of LLMs with the intention to continue the development of QA systems that are explainable, track provenance and adapt to changes in their knowledge-base [59].

## 2 Background

### 2.1 Zeolite Architectures and their Chemistry

Owing to their highly porous framework topologies, zeolites are significantly less dense than other silicate-based minerals (*e.g.* quartz). However, this aspect often increases their crystallographic complexity [45]. Standardising the description of these frameworks has been one of the main focal points of the *International Zeolite Association* [1], which has developed a variety of industry and research standards for zeolite chemistry, including codes of formally recognised zeolite materials, synthesis and characterisation references, among others. The association recognises over 250 topologically different zeolitic frameworks designated with three-letter codes. For instance, “Linde Type A” -LTA is one of the very commonly studied and described zeolite frameworks (see Figure 2). Although, one can build an LTA framework solely of Si as T atoms [14], this material is more commonly built of Si and Al atoms in equal ratios. In the latter case, owing to the Al-presence, the overall framework becomes formally negatively charged, and thus it attracts counter-cations in its pores. In the case of sodium cation incorporation, one forms zeolite material formulations of the type  $\text{[Na}_{12}(\text{H}_2\text{O})_{27}\text{[Al}_{12}\text{Si}_{12}\text{O}_{48}]}$ . The crystallographic unit cell of this zeolite is cubic ( $a = 24.61 \text{ \AA}$ ) with  $Fm\bar{3}c$  symmetry. The LTA framework has eight-member oxygen ring pores with a size of around  $4.3 \text{ \AA}$ .

Another example of a zeolite framework is FAU (Figure 2), whose three-letter code derives from the mineral faujasite. The naturally occurring faujasite exhibits a framework construction formula described as  $[\text{Al}_7\text{Si}_{17}\text{O}_{48}]^{7-}$ , which requires to be counterbalanced by cations. In the natural form, this can be based on  $\text{Na}^+$ ,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ , which collectively counter the charge, although their relative contributions can vary and may differ between samples. In synthetically formed FAU, the silica-to-alumina ratios may differ, while increased stability favours Si-rich frameworks. Furthermore, in synthetic FAU systems, the counter-cations can be similarly exchanged, leading to a plethora of different formulations. The unit cell of FAU zeolites is cubic with  $a = 24.65 \text{ \AA}$ ) and  $Fm\bar{3}m$  space-group symmetry.



**Figure 2:** Illustration of key concepts defining zeolite chemistry (top to bottom): CBUs describe the topology of frameworks, while based on a common framework, different formulations/materials can be derived. These materials differ in terms of the reported species/moieties they incorporate.

When comparing both framework types, one can notice particular similarities. First, the T-atoms virtually describe polyhedral cages that share polyhedral corners, edges and faces with their respective neighbours. These types of virtual framework building fragments are often referred to as composite building units and, in principle, can be discrete (e.g. rings and polyhedra) but also continuous (e.g. chains) [6, 49]. When examining LTA and FAU



frameworks, we notice that they both share structural arrangements, such as the sodalite cage made of 24 T atoms. This aspect is quite interesting as different fragments of the zeolite framework may be responsible for different functionalities. However, their description and existence provide a possibility for cross-structural comparisons. In addition to the composite building unit description, a more general description with mathematical tiling has evolved, which describes zeolitic topologies as three-dimensional structures made of polygonal faces that are commonly referred to as “Natural Building Units”, which do not necessarily need to be flat [11–13].

The zeolite crystal structures often display a plethora of species found in their cavities. These species may have entered the zeolite cavities through “post-synthetic” modifications such as ion exchange. Calcination is a process that normally removes internal species, but the charge balance is maintained through (partial) protonation. During the synthesis of zeolites, chemical species may play a role in directing the chemical outcome. However, their role may be conceptualised as a rigid templating effect, as it can be the case that a zeolitic framework can be synthesized in the presence of many different species [55]. Finally, complex zeolitic structures can also tightly incorporate complementary cluster materials that form simultaneously with the zeolite formation [40].

## 2.2 Crystallographic Information

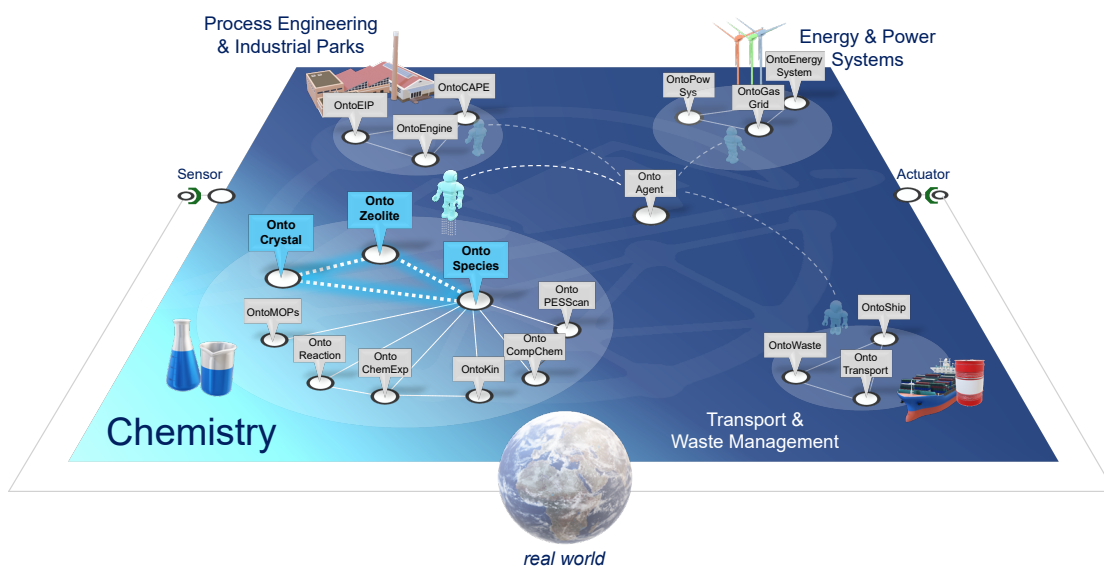
The CIF (Crystallographic Information File) is a structured text file format designed and maintained by the International Union of Crystallography (IUCr) for the storage of crystal structure data as well as information relating to the actual crystallographic measurement [10, 35]. The CIF contains different data blocks with array-like structuring covering information on atomic coordinates, lattice parameters, Miller indices, coordinate transformation matrices, Cromer-Mann scattering-factor coefficients *etc.* [70, 77] The core CIF dictionary is rich in terms of data names that enable convenient archiving and exchanging of raw and processed crystallographic data. This dictionary covers several thousand data properties; however, only 30 are sufficient to represent the crystallographic information involved in the virtual building of zeolitic models (see Figure S2 in SI). Many of the concepts (i.e. tags) covered by the core CIF and its related dictionaries relate to publication information, sample preparation, experimental conditions and techniques used, and audit and revision history, which are not involved in crystallographic model building but provide process information for reproducibility and data integrity purposes. In this regard, these concepts are useful for practical guidance on other integrated ontologies as part of wider laboratory digital twin [64].

Attempts to represent chemical crystallographic information with the help of the semantic web technologies have been reported [23]; however, the respective ontologies have not reached a maturity level to provide detailed representation for the complex query of crystals at the atomic level. The reason for this may be that to make meaningful queries, the data of the CIF has to undergo vector and matrix transformations, taking into consideration the overall crystallographic symmetry. In this work, we develop a new crystal structure describing ontology OntoCrystal, which includes classes that facilitate operations suitable for semantic storage of data as well as visualisation.

## 2.3 Digital Chemistry in The World Avatar

The World Avatar (TWA) is an open, dynamic world model built upon the semantic web stack (Figure 3). It encapsulates a comprehensive representation of diverse domains, including power and heat network optimisation, environmental monitoring, and climate resilience as demonstrated through the Climate Resilience Demonstrator (CReDo) project [5, 24]. Central to TWA’s functionality is its focus on chemicals and processes, underpinned by interlinked ontologies such as OntoSpecies, OntoKin, OntoCompChem, and OntoPESScan [26, 44, 51]. These ontologies provide a semantic framework for representing chemical species, reaction mechanisms, quantum chemistry calculations, and potential energy surface scans. Through its carefully designed interconnectivity, TWA promotes data interoperability and reduces ambiguity across previously isolated data silos [4, 28, 42, 54].

Semantic agents play a vital role within TWA, managing information flow and executing complex tasks. These agents perform essential functions, such as the calibration of kinetic mechanisms [7] and the automated design of metal-organic polyhedra (MOPs) based on inductive reasoning algorithms [32, 41]. To facilitate user interaction, TWA employs a question-answering system named “Marie”, which leverages advanced natural language processing to provide real-time responses [75, 87, 88]. The output agents that form the Marie functionality map natural language question to machine-readable SPARQL commands that retrieve the relevant information from TWA [42].



**Figure 3:** A selection of ontologies and their connectivity that have been integrated in TWA. OntoCrystal, OntoZeolite, and OntoSpecies are part of the digital chemistry domain.

## 3 Methodology

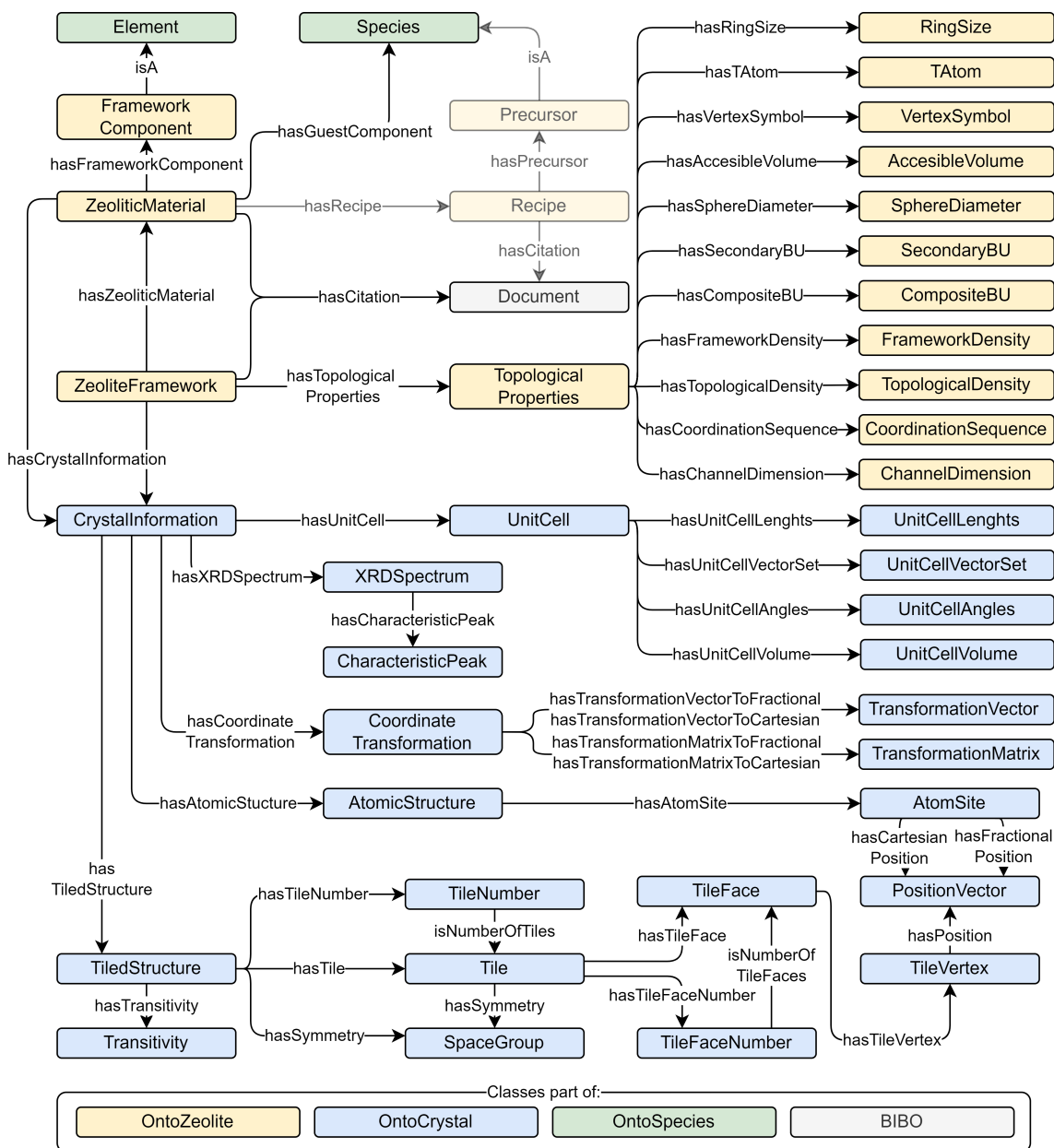
In the ontological context, the TBox (Terminological Box) organises and hierarchically categorises concepts while defining inter-domain associations through object properties. This can be represented through the help of description logic (see (see Section SI.4)). In contrast, the ABox (Assertional Box) leverages the TBox structure to instantiate these concepts with specific entities and their interrelations, as well as relevant data. Together, they enable precise data querying, individual entry access, and consistency checks [72]. The Hermit reasoning tool [33] checks the consistency of the TBox and ensures that the data types in the ABox align with the definitions provided in the TBox.

Prior to the creation of an ontology, we developed competency questions (see Section SI.2) to determine the scope of the ontology and ensure the ontological model captures complex domain interconnections. This section summarises the development of three critical ontologies: OntoZeolite, OntoCrystal, and OntoSpecies, each crucial for integrating domain-specific knowledge coherently.

### 3.1 OntoZeolite

The OntoZeolite ontology provides a structured framework that contextualises zeolites-related knowledge. This ontology introduces 26 classes, 26 object properties and 28 data properties. One of the central classes in this ontology is the `ZeoliteFramework`. This class is used to instantiate information about individual framework types (*e.g.* FAU, LTA, NAT *etc.*). A zeolite framework may be described separately or in combination through a set of topological properties. Thus, the `TopologicalProperties` class further connects to classes such as `AccessibleAreaPerCell`, `OccupiableVolume`, `FrameworkDensity`), (*e.g.*, `RingSizes` and other provide different information about the properties that define the frameworks.

The class `ZeoliteFramework` also connects to the class `ZeoliticMaterial`. The latter class is introduced to represent different zeolite instances that have been synthesised or discovered in nature. On practical grounds, for every zeolite material, we further represent the elements and their count involved in the description of the framework structure. In the ontology, this is being implemented through the class `FrameworkComponent`, which allows querying of materials based on elemental composition and relative compositions. Considering that within the `ZeoliticMaterial`, there can be different chemical species, they are represented as such through the class `Species` in the OntoSpecies ontology. As `ZeoliticMaterial` and `ZeoliteFramework` are crystalline in nature, they further connect to the class `CrystalInformation` defined by the OntoCrystal ontology. All zeolitic frameworks and materials are linked to the `Document` class. This class connects them to relevant bibliographic details using the BIBO ontology [21]. Considering the growing interest in digital exploration of the synthesis of new zeolite materials [58], our ontology also introduces a link between the `ZeoliticMaterial` and `Recipe` classes, followed by connections to `Precursor` chemicals and chemical `Species` for future studies.



**Figure 4:** Overview of the main classes, properties and interconnectivity between *OntoZeolite*, *OntoCrystal*, *OntoSpecies* and *BIBO* ontologies.

## 3.2 OntoCrystal

The *OntoCrystal* ontology provides a semantic representation of crystallographic data (see Figure 4). This ontology encompasses 18 classes, 43 object properties and 25 data properties. Physical properties with unit reuse concepts defined by the Ontology of Units of Measure (OM) version 2.0 [65, 66]. Crystal Information Files (CIF) allow and often use measured values with uncertainties, which are not currently supported by OM. For such data we introduced a new concept *MeasureWithUncertainty* in *OntoCrystal* (see SI, Figure S1 for further details).

The central class in the OntoCrystal ontology, `CrystalInformation`, is used to store fundamental crystallographic information and aggregates data from five key classes: `UnitCell`, `XRDSpectrum`, `AtomicStructure`, `CoordinateTransformation`, and `TiledStructure`. The `UnitCell` class provides metrics on unit cell dimensions, including lengths, vectors, angles, and volume. `AtomicStructure` details the arrangement of atoms within the crystal lattice. The atom site information consists of the atom type, the absolute and relative positions, and the site occupancy. The `CoordinateTransformation` class incorporates transformation vectors and matrices to convert relative within the unit cell to real Cartesian coordinates, and vice versa. The `XRDSpectrum` class models the X-ray Powder Diffraction spectrum, quantifying X-ray diffraction intensity across diffraction angles and is represented in a “ $2\theta$  plot”, which can be derived from experimental or simulated data. Apart from the full plot data represented as `PlotXY` this class stores the same information as a list of peaks. The `CharacteristicPeak` class is tailored for fingerprint analysis, facilitating the assessment of peak characteristics, including position, intensity, and width, critical for comparative crystallography. In most cases, the processed data in terms of `CharacteristicPeak` saves storage, and the full plot data is omitted in this case.

Natural tiling of space is a practical way of describing zeolite frameworks; however, its relevance is far more generally applicable to crystalline materials. The natural tiling involves the concept of tile which is also considered in by the CIF standards and described in a separate topology dictionary [10]. Thus as part of OntoCrystal, we included `TiledStructure` that defines the tiling patterns and includes the `Transitivity` class, which reflects on the uniformity and the description of the allowed transformations through symmetry operations. `TiledStructure` further connects to the classes `Tile`, `TileNumber` and `SpaceGroup` that define the geometric properties of tile faces, the count of tiles and the space groups associated with each tile configuration.

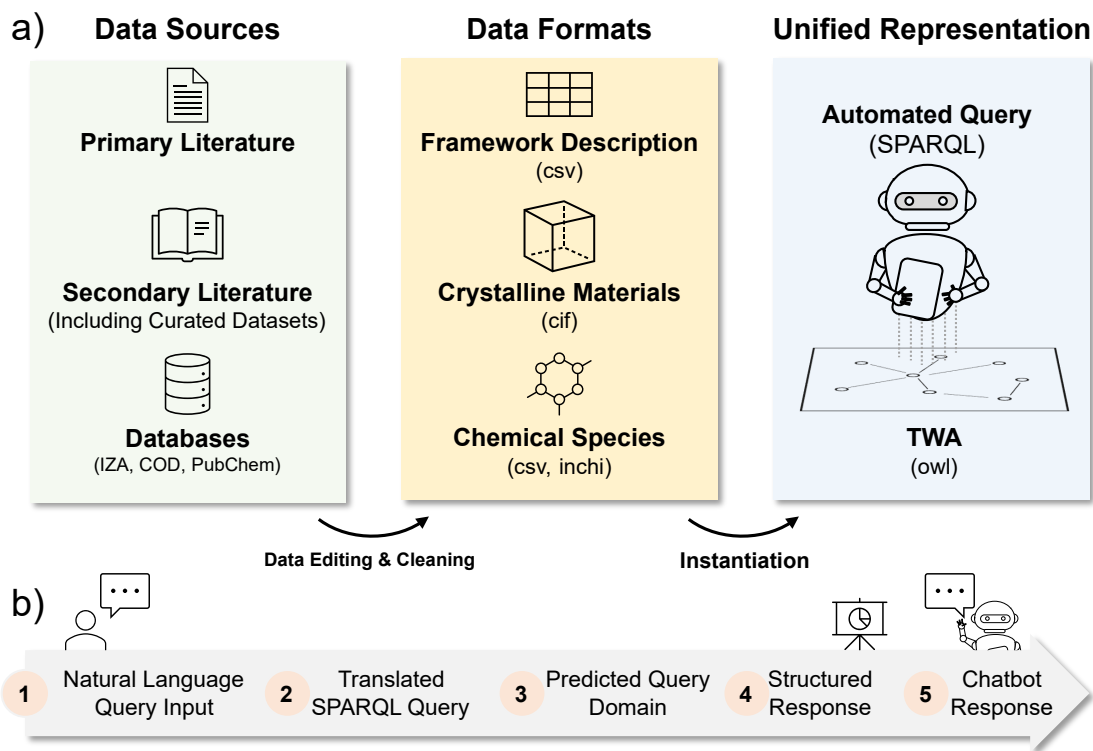
### 3.3 OntoSpecies

`OntoSpecies`, an integral ontology within the TWA framework, catalogues distinct chemical species and their properties, each assigned a unique Internationalized Resource Identifier (IRI) to ensure unambiguous identification [60]. This ontology works in tandem with `OntoZeolite` to facilitate the precise identification of chemical species in zeolite structures through `OntoSpecies` IRIs, thereby enabling detailed exploration of their interconnected properties. `OntoSpecies` is crucial in linking species to instances and concepts from other ontologies within the TWA chemistry domain. It also incorporates common cheminformatic identifiers such as InChI, InChIKey, CAS registry numbers, PubChem CID, and SMILES, which are used for retrieving external information. The molecular geometry is meticulously documented within the ontology, making the data usable for quantum chemical calculations, with each bond and atom distinctly identified by an IRI. The `OntoSpecies` ontology encompasses a broad spectrum of chemical and physical properties, classifications, applications, and spectral data for each species. It includes detailed provenance and attribution metadata to ensure the reliability and traceability of the data. Most of the chemical species information is sourced from a variety of open chemical databases, rendering `OntoSpecies` as a unifying ontology for chemical informatics.

### 3.4 Data Curation

Information on zeolites and there has been guided by the IZA structural dataset in conjunction with zeolite framework and material descriptions published as original research [3, 6]. From the original literature, we have acquired information on mineralogical and synthetically reported zeolites, which includes their chemical formula, crystallographic information and relation/incorporation of chemical species/counterions in its porous structure. Additional information on zeolite materials, their chemical formulas, relation to crystallographic systems, and their bibliographic information were sourced from previously published and peer-reviewed datasets [83, 85]. Manual cross-checking of papers was required to confirm the presence of chemical species/ions, and further collection on the properties of these chemical species and ions was performed through programmatic queries from the PubChem database. In a few instances, PubChem info was absent (*e.g.* for cluster and organometallic structures), and thus such instances were added manually.

The original data were derived from various file formats, including CSV, CIF, JSON, BIB, and TXT, among others. Following this, as outlined in our workflow (see Figure 5.a), we augmented, corrected, and supplemented missing data as necessary. For XRD spectra, we extracted the  $2\theta$  positions and their relative intensities, preparing them for instantiation. All data formats were augmented to produce an OWL ABox, which was subsequently uploaded to our knowledge graph. Comprehensive details on the data curation process are available in the supporting information (see SI.1 for more details).



**Figure 5:** Overview of a) the data curating and processing workflow; b) processing of natural language queries on TWA-Marie interface.

### 3.5 TWA Integrated Query Interface

To facilitate the exploration of zeolite chemistry, a user interface was developed, enabling efficient interaction with data on zeolite properties (see Figure 5.b). This interface provides both field-based and natural language search options, and it is currently available through the TWA-Marie webpage, equipped with plentiful examples across various chemistry domains (see <https://theworldavatar.io/demos/marie/> for more detail).

The structured or field-based search feature related to zeolitic frameworks enables cross-structural comparison by plotting numerical data of over twenty different properties. This built-in functionality comes with the calculation of correlation coefficient and colour mapping based on a third property. Additionally, frameworks and material instances can be queried using pre-defined search fields. In the case of zeolite frameworks, users can query framework information based on X-ray diffraction (XRD) peak positions and their relative intensities, unit cell parameters, different forms of densities and building unit features describing the framework topology. Meanwhile, zeolitic materials can be retrieved based on their formula, elements that form the framework, and non-framework species/ions. As crystallographic information and academic literature are associated with the zeolitic material instances, they can also be queried using unit cell parameters and DOI numbers.

Unstructured or natural language search allows users to submit a query in natural language without locating specific input fields; users then obtain responses in both tabular and human-friendly textual formats. This is achieved by applying our previously developed method that supports our question-answering system for combustion kinetics [61]. Specifically, we performed multi-task fine-tuning on the pre-trained language model FLan-T5 for natural language-to-SPARQL translation and domain classification tasks. At test time, the model runs two inference tasks: translating natural language input into a corresponding SPARQL query and predicting TWA domain for SPARQL execution to retrieve desired information (see Section SI.2 in the Appendix for a detailed process breakdown).

## 4 Results and Discussion

In this section, we provide an overview of the zeolite and crystallographic information within the context of TWA. We demonstrate the semantic structuring and interconnection of zeolitic, crystallographic, and species data through SPARQL queries. These queries, developed with the ontological structure in mind, enable programmatic searches. However, crafting queries may not always be straightforward. Therefore, template queries can be developed and deployed for advanced searches, either through a web interface or within a question-answering system.

### 4.1 Overview on TWA Zeolitic Instances

After instantiation of zeolite framework, material, species and crystallographic information, TWA provides coverage of 251 zeolites, over two thousand zeolite materials where the majority are supported by crystallographic information. In Figure 6, we attempt to

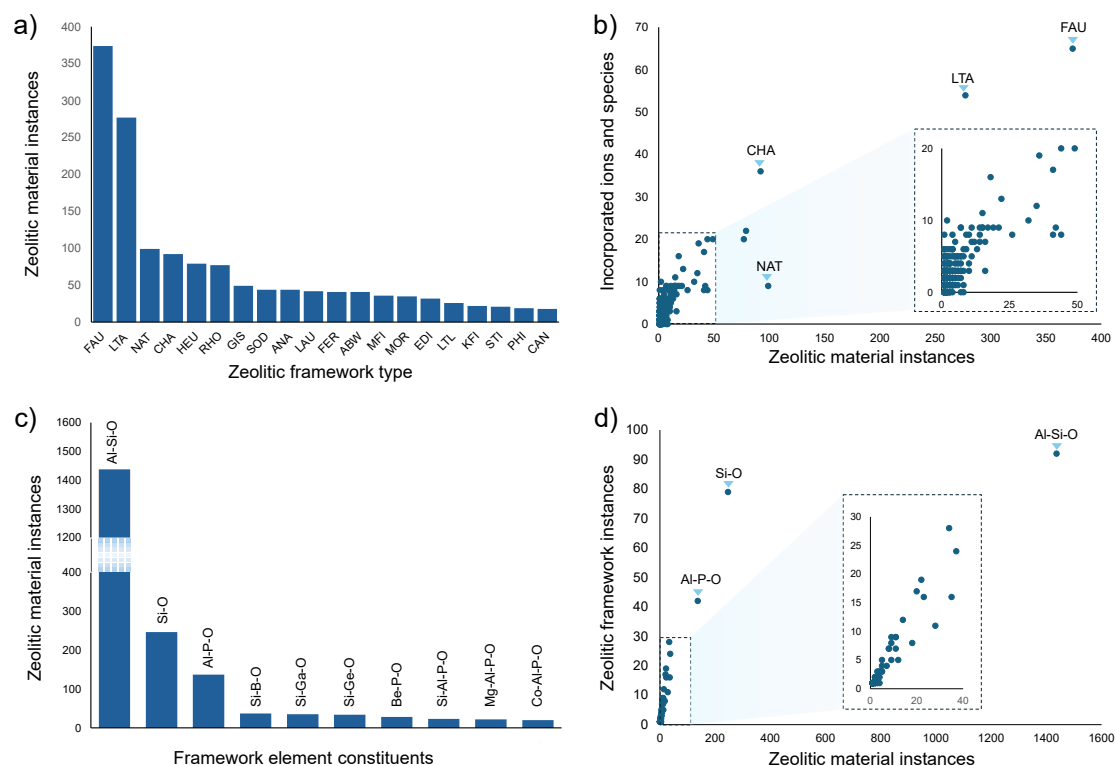
analyse the available data on how the different zeolitic instances are distributed across framework types and what sort of species/ions they incorporate.

The top 10 zeolitic frameworks—namely FAU, LTA, NAT, CHA, HEU, RHO, GIS, SOD, ANA, and LAU—encompass a total of 1177 instances, as demonstrated in Figure 6.a. This high instance density per framework indicates that a relatively small number of zeolitic frameworks are the focus of a significant portion of scientific inquiry and analysis within the field. The FAU framework, in particular, registers the highest occurrence with 374 instances, followed by the LTA framework with 277 instances and NAT and CHA frameworks with 99 and 92 instances, respectively. The aggregation of these instances among the top frameworks indicates their importance and the scientific community's interest in their unique properties and applications.

Subfigure 6.b presents a scatter plot that examines the correlation between the number of reported material instances of zeolite frameworks and the diversity of incorporated ions and species. While the data points predominantly cluster near the origin, indicating a prevalent trend of limited incorporation diversity across most frameworks, a few notable exceptions emerge. The frameworks of FAU, LTA, CHA, and NAT distinguish themselves not only through a higher count of material instances—374, 277, 92, and 99, respectively—but also through their considerable diversity of ions and species, with FAU, LTA, CHA, and NAT having 65, 54, 36, and 9 unique guest components, respectively. Together with HEU, RHO, and GIS frameworks, these seven types account for over 1000 material instances, demonstrating their importance and potential for structural and chemical adaptability. Limitations in terms of the diversity of incorporated species are obvious in the case of the NAT framework. This frequently studied has been found to form mainly in the presence of sodium cations, which explains the low variety of incorporated species. From the collected information, an overwhelming majority of zeolite framework types—approximately 92.3% have been associated with less than 25 material instances and fewer than ten different ions or species. This stark contrast indicates that a small minority of zeolite frameworks are associated with most of the reported zeolite materials and incorporated species.

Within our knowledge graph, there are 73 distinct sets of framework-building elements. A significant majority of these, comprising 1437 instances, consist of aluminium and silicon, as illustrated in Figure 6.c. This prevalence aligns with the common definition of zeolites as hydrated aluminosilicates often containing sodium, potassium, calcium, and other cations. Correspondingly, aluminosilicates dominate within the largest set of framework topologies (92 topologies), as depicted in Figure 6.d. Following aluminosilicate zeolites, purely silicate-based frameworks are the second most represented, with 247 instances across 79 topologies. Aluminophosphates also feature prominently, with 137 material instances spread over 42 topologies. Beyond these three prevalent material types, our knowledge graph encompasses a variety of structures composed of different elemental combinations.





**Figure 6:** Overview of the zeolitic instances from the TWA: (a) Bar chart displaying zeolitic framework types alongside their quantity of material instances. (b) A scatter plot shows the number of material instances for each framework type versus the diversity of incorporated species. (c) Bar chart detailing the frequency of the 10 most prevalent framework-building elements and their various combinations. (d) Scatter plot presenting the quantity of zeolitic material instances in relation to the range of elemental combinations within framework types.

## 4.2 Custom SPARQL-based Requests

SPARQL is an RDF query language designed to retrieve semantically structured data. It is paired with Blazegraph [2], an open-source triplestore that serves as the main graph database infrastructure of TWA. Understanding the ontological structure enables the crafting and execution of customised SPARQL queries over Blazegraph for programmatic data retrieval. Examples of such queries, specifically for accessing crystallographic information via our web-hosted [TWA - Blazegraph](#), are documented in the supporting information of this work (Section [SI.3](#)). Figure 7 illustrates a query retrieving chemical information about species within zeolites. To extract the required data, the system navigates the graph, starting from the specified zeolite framework to associated material instances and then to the interconnected species IRI, before retrieving details about those chemical species (*e.g.*, molecular weight). Typically, chemists with domain expertise extract such information manually through cognitive processes. This case demonstrates how TWA can perform cognitive-like tasks, by navigating its knowledge graph [42].

```

PREFIX zeo: <http://www.theworldavatar.com/kg/ontozeolite/>
PREFIX os: <http://www.theworldavatar.com/ontology/ontospecies/OntoSpecies.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?ZeoliteName (SAMPLE(?IUPAC) AS ?GuestIUPACName) ?GuestFormula ?MolecularWeight
↪ (SAMPLE(?BPValue) AS ?BoilingPoint_value) (SAMPLE(?BPUnit) AS ?BoilingPoint_unit)
WHERE {
  ?Framework zeo:hasFrameworkCode "LAU" ;
             zeo:hasZeoliticMaterial ?Material .
  ?Material  rdf:type zeo:ZeoliticMaterial ;
             os:name ?ZeoliteName ;
             zeo:hasGuestCompound ?Guest .
  SERVICE <https://theworldavatar.io/chemistry/blazegraph/namespace/ontospecies> {
    ?Guest  rdfs:label ?GuestFormula ;
           os:hasIUPACName/os:value ?IUPAC ;
           os:hasMolecularWeight/os:value ?MolecularWeight .
  }
  OPTIONAL {
    ?Guest  os:hasBoilingPoint ?BPNode .
    ?BPNode os:value ?BPValue ;
           os:unit/rdfs:label ?BPUnit ;
           os:hasReferenceState/os:value ?value .
  }
}
GROUP BY ?ZeoliteName ?GuestFormula ?MolecularWeight ORDER BY ?ZeoliteName

```

**Figure 7:** Example of a SPARQL query that retrieves information cross Zeolitic framework, Zeolitic material to molecule species. Example output is Figure S6

## 4.3 Web-assisted Data Exploration

TWA is the first instance of semantically-assisted machine-to-machine communication [50], however, to enable humans to interact with TWA, tools to overcome the human-machine barrier are needed. In this section, we showcase examples where SPARQL queries are automatically drafted based on user input. We first show a query of properties across zeolitic frameworks, which are then illustrated using the built-in plotting and correlation tools property correlations (Subsection 4.3.1). Next, we show a query of frameworks and material instances based on pre-drafted inputs 4.3.2 and this type of search is also extended to finding structural models based on powder XRD peak positioning (Subsection 4.3.3). Finally, we cover examples of querying TWA with the help of natural language processing (subsection 4.3.4).

### 4.3.1 Queries for Cross-Framework Comparisons

The *Zeolite Explorer* tool facilitates the identification of overarching trends in zeolite frameworks by allowing users to input specific framework parameters. Upon specification, the system populates a predefined SPARQL query, which retrieves the corresponding data. This data is then visually represented in a color-coded two-dimensional plot. This section discusses the range of parameters that users can explore, highlighting that not all parameters exhibit significant correlations. Conversely, properties such as the number of tile edges and ring member sizes generally demonstrate strong correlations due to their shared structural roles in defining framework characteristics.

An interesting aspect of zeolite materials is the distinction between accessible and occupiable areas. Although zeolites can have large cavity cages, their accessibility is often limited by the small size of the channels leading to them. For most structures, the occupiable area generally correlates linearly with the accessible area, demonstrated by a correlation coefficient of 0.83, as shown in Fig.8.a and Fig.8.b. These plots employ different colour schemes to highlight the largest and smallest ring sizes, respectively, but both illustrate the same underlying data relationship. This pattern is exemplified by the sodalite framework (SOD), which does not show any accessible area in Fig.8.a, despite its large cavity sizes. In contrast, zeolites with highly accessible areas over 2500 m<sup>3</sup>/g often have ring sizes exceeding 10 members but also include some of the smallest rings, as seen in Fig.8.b. This variability can be attributed to the cage structures resembling truncated polyhedra, where truncation forms openings of various sizes, enhancing internal accessibility.

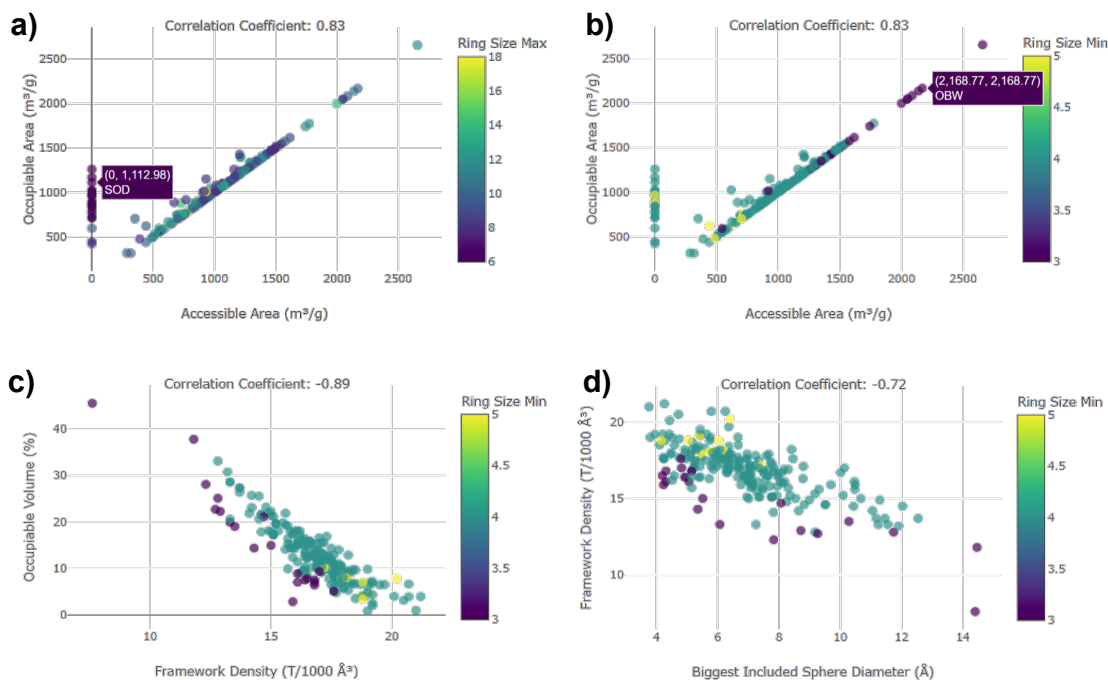
In Fig.8.c, a clear trend is observed where the occupiable volume decreases as the framework density increases (correlation coefficient = -0.89), indicating a strong inverse relationship. Framework density, defined as the number of tetrahedral atoms per 1000 cubic angstroms (Å<sup>3</sup>), is inversely correlated with pore size. This relationship illustrates that denser structures, characterised by smaller pores, offer less available cavity space. Further analysis highlights a correlation between ring size and framework density: structures with ring sizes of 3 are generally less dense, featuring more expansive cavities, whereas zeolites with a minimum ring size of 5 are among the most densely packed, leading to significantly lower occupiable densities (correlation coefficient = -0.72). This pattern is also supported by the largest included sphere diameter, indicating that the least framework density is typically found in zeolites with a minimum ring size of 3, as depicted in Fig.8.d.

### 4.3.2 Structured Query of Framework Properties

The graph-based structuring of knowledge in zeolitic materials promotes efficient navigation through interconnected information, utilizing common edges and nodes to link related entities. Our ontology structures, illustrated in Figure 4, facilitate straightforward transitions from species-level data to zeolite frameworks and subsequently to their crystalline properties. This architecture supports queries such as: “*find property X of zeolite(s) Z belonging to a given framework F*”. Further, as shown in Figure 9, users can query: “*Find all Frameworks F that have properties X satisfying given conditions*”. These queries return detailed information about zeolite frameworks, including crystalline structure and porosity. Access to these properties is provided through the *Advanced Search* function of the web interface, which offers input fields for specifying values, value ranges, or strings related to various material properties. Upon input submission, the backend processes the data into one or several SPARQL queries, each producing a list of results. The final output, an intersection of these lists, ensures comprehensive and accurate retrieval of data.

### 4.3.3 Structured Identification of Reference XRD Powder Patterns

Reference powder X-ray diffraction (XRD) spectra comprise published spectral data of materials that have undergone rigorous verification processes. In the field of zeolite chemistry, reference spectra are available for the majority of framework compounds [76]. These

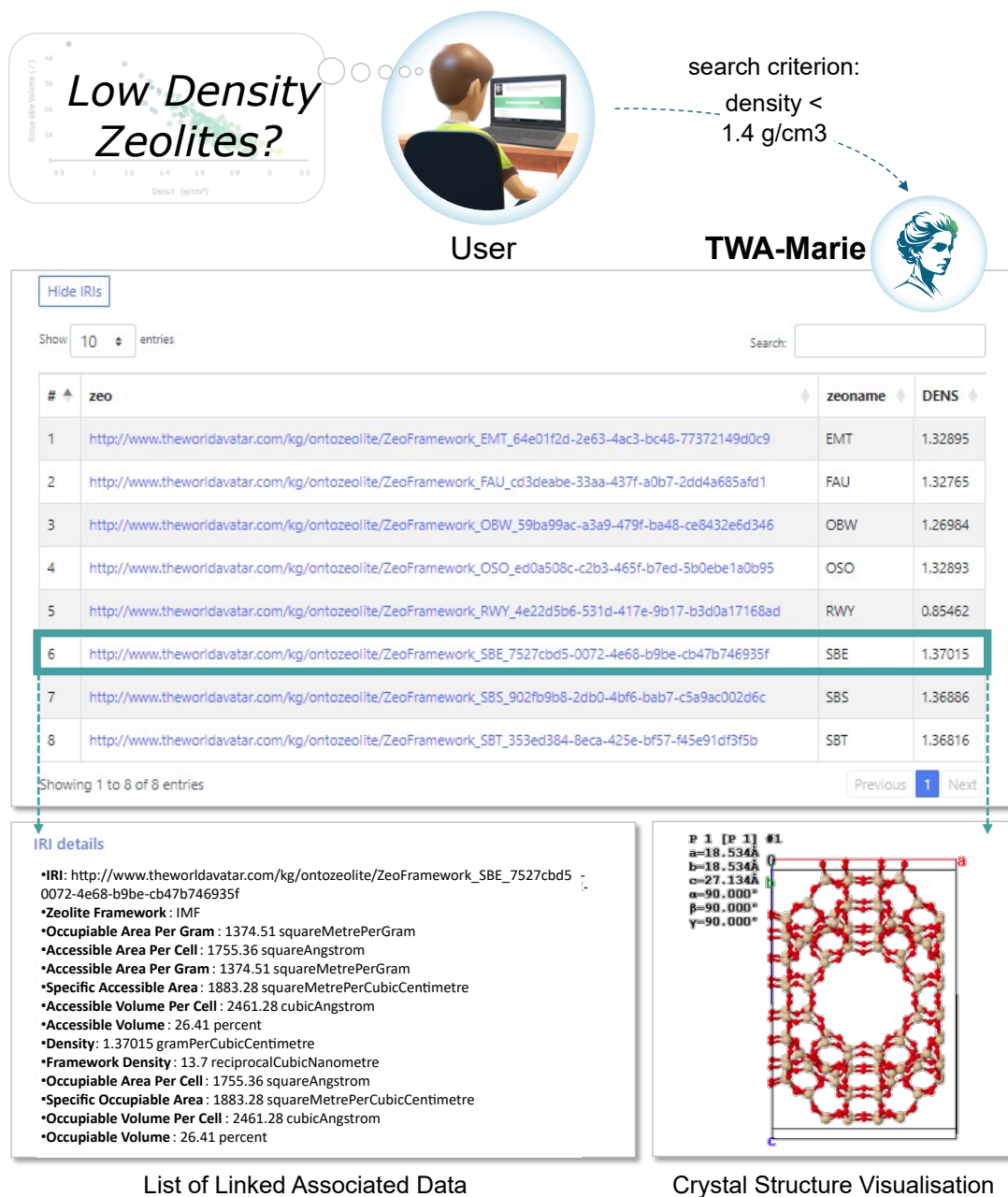


**Figure 8:** Correlations and properties of zeolite frameworks. (a) accessible area vs occupiable area, colouring: ring sizes max; (b) accessible area vs occupiable area, colouring: ring sizes min; (c) framework density vs occupiable volume, colouring ring size min; (d) biggest included sphere diameter vs framework density, colouring: ring size min.

spectra are instrumental in spectral blueprinting analysis, a methodology employed by researchers to ascertain whether a newly synthesized zeolite material matches an existing framework. This blueprinting process involves a detailed comparison of characteristic diffraction peaks. Blueprinting of XRD spectra has the opportunity to be automated, thus enhancing the accuracy and efficiency of the material verification process.

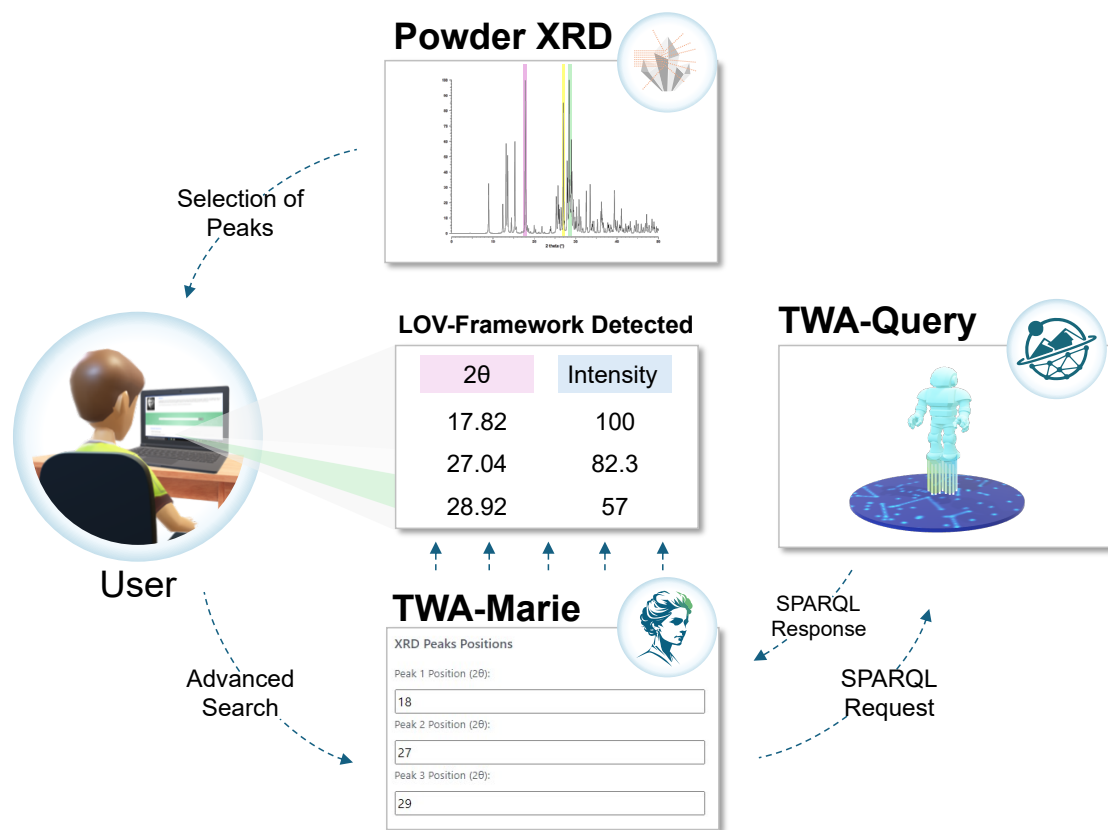
In our knowledge model, the XRD powder data is linked to zeolitic frameworks. However, signal positions and relative intensities are crucial for the fingerprint identification of structures, and thus, we effectively use them to query and predict XRD plots based on user input. The whole operation involves SPARQL queries, which retrieve this data, compare it with the user’s input, and suggest a framework type that has been identified through the fingerprinting method. The templated SPARQL queries are adjusted to essentially respond to the question “find frameworks F that have peaks of relative height at least  $P_I$  near a given position  $P_{2\theta}$ ”. In our SPARQL template, we have provided the opportunity for up to three characteristic peaks given a position and intensity. The default width and the cut-off intensity used in the templated queries on the backend are  $0.5^\circ 2\theta$  and 50%, respectively. Examples of this query can be when a user inputs three  $2\theta$  positions: 18, 27, and 29. The query system might suggest that the closest match is with the LOV framework, where the positions are 17.82, 27.04, and 28.92 (Fig.10).

In contrast to recent studies employing machine learning for the comparison of XRD



**Figure 9:** An example of web-assisted SPARQL query where a user specifies a parameter (e.g. low-density zeolitic frameworks and only through the query of TWA complete information obtained as well as a structural projection of the zeolitic material.

powder spectra in Metal-Organic Frameworks (MOFs) [79], the current approach offers an expandable knowledge base and relies on high-quality reference data.



**Figure 10:** Schematic illustration of search of matching framework types based on XRD diffraction characteristic peaks.

#### 4.3.4 Question-Answering based using the Marie System

The natural language interface of Marie enables the retrieval of chemical information across networks of interlinked data via a single search entry point. We demonstrate the capability of this feature with a run-through of the steps involved in the natural language processing pipeline, as seen in Fig. 11. The user first inputs a natural language query asking for a list of zeolite materials made of Ge and O only. When the user presses *Enter* or clicks the button with the magnifying glass icon, the system performs two operations: translating the user input into a SPARQL query and identifying which knowledge domain to query; the results of this step are shown below the input field, with the SPARQL query displayed on the left and predicted domain, which is *OntoZeolite*, on the right. The SPARQL query is then executed against the target knowledge graph to obtain the requested information in a tabular format, showing the chemical formulae of the zeolitic materials asked for by the user. To further enhance the user experience, this structured data, together with the input question, is passed to OpenAI’s chat completions API to formulate a concise, human-friendly chatbot response that directly answers the user query [57].

In evaluating the performance of TWA-Marie, commercial ChatGPT 4, and Gemini Advanced within zeolite chemistry, notable differences in accuracy, detail, and reliability are

observed (see Section as shown in Section SI.2 for more details). TWA-Marie combines knowledge graph information with a large language model to deliver precise and reliable information substantiated by direct IRI and DOI links. For instance, inquiries regarding the reported unit cell parameters of specific zeolite framework types such as ABW, AHT, and LAU consistently receive accurate responses. In contrast, ChatGPT demonstrates inconsistent accuracy, occasionally providing incorrect or hallucinated data, including misidentifying the crystal system of zeolite ABW or conflating the LAU framework with LTA. Similarly, Gemini Advanced’s responses often contain inaccuracies or information irrelevant to the queries posed, like in cases where it is asked about zeolites reported to include pyridine within their frameworks. These discrepancies highlight the superiority of TWA-Marie’s approach, integrating a knowledge graph with a large language model to provide data-driven and verifiable responses.

**1 Natural Language Query Input:** The user asks for a list of zeolite materials with framework made of Ge

Common units with special characters: `<code>?Ge?</code>`

Find zeolites with framework made of Ge and O only

**2 Translated SPARQL query:** The input query in natural language is translated to SPARQL

```
SELECT ?Material WHERE { ?Material zeo:hasFrameworkComponentOnly/rdfs:label "Ge". ?Material zeo:hasFrameworkComponentOnly/rdfs:label "O". }
```

**3 Predicted query domain:** The domain is predicted to be OntoZeolite

Predicted query domain: ontozeolite  
 Translation latency: 2.70s.  
 SPARQL query execution latency: 4.28s.  
 Chatbot response latency: 2.15s.

**4 Structured Response:** The predicted SPARQL query is executed against the OntoZeolite knowledge graph, returning the formula of zeolite materials in a tabular format.

#	ChemicalFormula
1	(DMA)2(H2O)2 [Ge20O40]
2	(CID12127913)F [Ge40O80]
3	(DABCO)20(H2O)20 [Ge40O80]
4	(HM+2)2F4 [Ge40O80]
5	(TrMA)6(H2O)4.5 [Ge32O64]

**5 Chatbot Response:** A large language model processes both the data from the knowledge graph and the input query to output a human-friendly list of zeolite materials.

**Marie's response**

The retrieved data contains information about several zeolites with frameworks made of Ge and O only. Here are the chemical formulas of those zeolites:

1. |(DMA)2(H2O)2|[Ge20O40]
2. |(CID12127913)F|[Ge40O80]
3. |(DABCO)20(H2O)20|[Ge40O80]
4. |(HM+2)2F4|[Ge40O80]
5. |(TrMA)6(H2O)4.5|[Ge32O64]

These zeolites consist solely of the elements Ge (Germanium) and O (Oxygen) in their frameworks.

```
SELECT DISTINCT ?Material ?ChemicalFormula ?Element1 ?Element2 WHERE {
  ?Material zeo:hasChemicalFormula ?ChemicalFormula .
  ?Material zeo:hasFrameworkComponent ?Element1 .
  ?Material zeo:hasFrameworkComponent ?Element2 .
  FILTER NOT EXISTS {
    ?Material zeo:hasFrameworkComponent ?ElementExclude .
    FILTER ( ?ElementExclude NOT IN ( ?Element1, ?Element2, ?Oxygen ) )
  }
  SERVICE <> {
    ?Element1 (os:hasElementSymbol|os:hasElementName)/os:value "Ge" .
    ?Element2 (os:hasElementSymbol|os:hasElementName)/os:value "O" .
    ?Oxygen os:hasElementSymbol/os:value "O" .
  }
}
```

**Figure 11:** The user interface for natural language search, with a breakdown of the processing steps involved.

## 5 Summary and conclusions

In this paper, we have detailed a semantic integration of concepts from zeolite chemistry with those of crystalline materials, alongside a focus on chemical species. This

integration has been achieved through the curation of chemical, crystallographic, and zeolite data, formatted according to an established ontological framework. We populated a comprehensive knowledge graph within the broader TWA model, covering frameworks associated with over two thousand zeolite materials and more than one thousand crystallographic structures linked to over 200 chemical species. This integration ensures that the chemical information becomes machine-actionable, enhancing the efficiency and precision of data queries and retrieval processes. This compatibility enhances the accessibility and actionability of complex chemical data by facilitating its delivery in precise, natural language. Moreover, the combination of the knowledge graph approach with LLM showed a distinct advantage over systems that depend solely on LLMs, which are prone to inaccuracies and data “hallucinations”.

Considering the relevance and need for programmatic study of crystalline information in drug design and materials engineering [16, 56], the presently reported ontological approach provides a promising alternative for crystallographic queries in the near future. In the context of zeolite research, TWA has the capacity to grow and further interconnect mechanical properties, gas adsorption characteristics, and synthetic data for known zeolites [15, 25, 39, 58]. The integration of diverse but interrelated chemical concepts enables tackling complex multicomponent chemical systems such as surface chemistry, reticular chemistry, and supramolecular chemistry [42]. This approach offers significant potential for interoperability within complex chemical material systems, thereby motivating continued exploration and detailed characterisation of these systems.

## Author Contributions

AK and MK conceptualized the study. AK authored the original manuscript, which was edited and revised by all authors. PR developed the ontologies with contributions from FF, LP, and AK. DT trained the question-answering system. LP developed the web interface for queries. SG validated the agent instantiation. Data curation was managed by A.K., LP, and PR. MK acquired research funding.

## Acknowledgements

This research was supported by the National Research Foundation, Prime Minister’s Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any arising Author Accepted Manuscript version.



## References

- [1] International Zeolite Association (IZA). <https://www.iza-online.org/>. Accessed: April 25, 2024.
- [2] Blazegraph™ DB, 2024. URL <https://blazegraph.com/>. Last accessed: 2024-04-12.
- [3] Database of zeolite structures. re3data.org - Registry of Research Data Repositories, 2024. doi: [10.17616/R3HS6N](https://doi.org/10.17616/R3HS6N).
- [4] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft. Universal digital twin – a dynamic knowledge graph. *Data-Centric Eng.*, 2:e14, 2021. doi:[10.1017/dce.2021.10](https://doi.org/10.1017/dce.2021.10).
- [5] J. Akroyd, A. Bhave, G. Brownbridge, E. Christou, M. D. Hillman, M. Hofmeister, M. Kraft, J. Lai, K. F. Lee, S. Mosbach, D. Nurkowski, and O. Parry. *Building a Cross-Sector Digital Twin*. Centre for Digital Built Britain, 2022. doi:[10.17863/CAM.81779](https://doi.org/10.17863/CAM.81779). Centre for Digital Built Britain.
- [6] N. A. Anurova, V. A. Blatov, G. D. Ilyushin, and D. M. Proserpio. Natural tilings for zeolite-type frameworks. *J. Phys. Chem. C*, 114(22):10160–10170, 2010. doi:[10.1021/jp1030027](https://doi.org/10.1021/jp1030027).
- [7] J. Bai, R. Geeson, F. Farazi, S. Mosbach, J. Akroyd, E. Bringley, and M. Kraft. Automated calibration of a poly(oxyethylene) dimethyl ether oxidation mechanism using knowledge. *J. Chem. Inf. Model.*, 61(4):1701–1717, 2021. doi:[10.1021/acs.jcim.0c01322](https://doi.org/10.1021/acs.jcim.0c01322).
- [8] J. Bai, S. Mosbach, C. J. Taylor, D. Karan, K. F. Lee, S. D. Rihm, J. Akroyd, A. A. Lapkin, and M. Kraft. A dynamic knowledge graph approach to distributed self-driving laboratories. *Nat. Commun.*, 15(1):462, 2024. doi:[10.1038/s41467-023-44599-9](https://doi.org/10.1038/s41467-023-44599-9).
- [9] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Sci. Am.*, 284(5): 34–43, 2001. doi:[10.1038/scientificamerican0501-34](https://doi.org/10.1038/scientificamerican0501-34).
- [10] H. J. Bernstein, J. C. Bollinger, I. D. Brown, S. Gražulis, J. R. Hester, B. McMahon, N. Spadaccini, J. D. Westbrook, and S. P. Westrip. Specification of the Crystallographic Information File format, version 2.0. *J. Appl. Crystallogr.*, 49(1):277–284, Feb 2016. doi:[10.1107/S1600576715021871](https://doi.org/10.1107/S1600576715021871).
- [11] V. A. Blatov, O. Delgado-Friedrichs, M. O’Keeffe, and D. M. Proserpio. Three-periodic nets and tilings: natural tilings for nets. *Acta Crystallogr. A*, 63(5):418–425, Sep 2007. doi:[10.1107/S0108767307038287](https://doi.org/10.1107/S0108767307038287).
- [12] V. A. Blatov, G. D. Ilyushin, and D. M. Proserpio. The zeolite conundrum: Why are there so many hypothetical zeolites and so few observed? a possible answer from the zeolite-type frameworks perceived as packings of tiles. *Chem. Mater.*, 25(3): 412–424, 2013. doi:[10.1021/cm303528u](https://doi.org/10.1021/cm303528u).

- [13] V. A. Blatov, A. P. Shevchenko, and D. M. Proserpio. Applied topological analysis of crystal structures with the program package ToposPro. *Cryst. Growth Des.*, 14(7):3576–3586, 2014. doi:10.1021/cg500498k.
- [14] B. W. Boal, J. E. Schmidt, M. A. Deimund, M. W. Deem, L. M. Henling, S. K. Brand, S. I. Zones, and M. E. Davis. Facile synthesis and catalysis of pure-silica and heteroatom LTA. *Chem. Mater.*, 27(22):7774–7779, 2015. doi:10.1021/acs.chemmater.5b03579.
- [15] N. S. Bobbitt, K. Shi, B. J. Bucior, H. Chen, N. Tracy-Amoroso, Z. Li, Y. Sun, J. H. Merlin, J. I. Siepmann, D. W. Siderius, et al. MOFX-DB: An online database of computational adsorption data for nanoporous materials. *J. Chem. Eng. Data*, 68(2):483–498, 2023. doi:10.1021/acs.jced.2c00583.
- [16] M. J. Bryant, S. N. Black, H. Blade, R. Docherty, A. G. Maloney, and S. C. Taylor. The CSD drug subset: The changing chemistry and crystallography of small molecule pharmaceuticals. *J. Pharm. Sci.*, 108(5):1655–1662, 2019. doi:10.1016/j.xphs.2018.12.011.
- [17] A. W. Burton and S. I. Zones. Organic molecules in zeolite synthesis: their preparation and structure-directing effects. *Stud. Surf. Sci. Catal.*, 168:137–179, 2007. doi:10.1016/S0167-2991(07)80793-2.
- [18] C. M. Castro Nascimento and A. S. Pimentel. Do large language models understand chemistry? a conversation with chatgpt. *J. Chem. Inf. Model.*, 63(6):1649–1655, 2023. doi:10.1021/acs.jcim.3c00285.
- [19] W. Chaikittisilp. *Data-Driven Approach for Rational Synthesis of Zeolites and Other Nanoporous Materials*, chapter 9, pages 233–250. John Wiley & Sons, Ltd, 2023. ISBN 9781119819783. doi:10.1002/9781119819783.ch9.
- [20] C. Chizallet, C. Bouchy, K. Larmier, and G. Pirngruber. Molecular views on mechanisms of brønsted acid-catalyzed reactions in zeolites. *Chem. Rev.*, 123(9):6107–6196, 2023. doi:10.1021/acs.chemrev.2c00896.
- [21] DCMI Usage Board. Bibliographic ontology (bibo) in rdf. Maintainer: DCMI Usage Board (contact: Bruce d’Arcus), 2016-05-11. URL <https://www.dublincore.org/specifications/bibo/bibo/bibo.rdf.xml>. Creators: Bruce D’Arcus, Frédéric Giasson.
- [22] J. Deb, L. Saikia, K. D. Dihingia, and G. N. Sastry. Chatgpt in the material design: Selected case studies to assess the potential of chatgpt. *J. Chem. Inf. Model.*, 2024. doi:10.1021/acs.jcim.3c01702.
- [23] G. Deepak, Z. Gulzar, and A. A. Leema. An intelligent system for modeling and evaluation of domain ontologies for crystallography as a prospective domain with a focus on their retrieval. *Comput. Electr. Eng.*, 96:107604, 2021. ISSN 0045-7906. doi:10.1016/j.compeleceng.2021.107604.

- [24] Digital Twin Hub. Climate resilience demonstrator. <https://digitaltwinhub.co.uk/credo/credo/>, 2023. Accessed: March 5, 2024.
- [25] J. D. Evans and F.-X. Coudert. Predicting the mechanical properties of zeolite frameworks by machine learning. *Chem. of Mat.*, 29(18):7833–7839, 2017. doi:10.1021/acs.chemmater.7b02532.
- [26] F. Farazi, J. Akroyd, S. Mosbach, P. Buerger, D. Nurkowski, M. Salamanca, and M. Kraft. Ontokin: An ontology for chemical kinetic reaction mechanisms. *J. Chem. Inf. Model.*, 60(1):108–120, 2020. doi:10.1021/acs.jcim.9b00960.
- [27] F. Farazi, N. B. Krdzavac, J. Akroyd, S. Mosbach, A. Menon, D. Nurkowski, and M. Kraft. Linking reaction mechanisms and quantum chemistry: An ontological approach. *Comput. Chem. Eng.*, 137(9):106813, 2020. doi:10.1016/j.compchemeng.2020.106813.
- [28] F. Farazi, M. Salamanca, S. Mosbach, J. Akroyd, A. Eibeck, L. K. Aditya, A. Chadzynski, K. Pan, X. Zhou, S. Zhang, M. Q. Lim, and M. Kraft. Knowledge graph approach to combustion chemistry and interoperability. *ACS Omega*, 5(29):18342–18348, 2020. doi:10.1021/acsomega.0c02055.
- [29] E. M. Flanigen. Chapter 2 zeolites and molecular sieves an historical perspective. In H. van Bekkum, E. Flanigen, and J. Jansen, editors, *Introduction to Zeolite Science and Practice*, volume 58 of *Studies in Surface Science and Catalysis*, pages 13–34. Elsevier, 1991. doi:10.1016/S0167-2991(08)63599-5.
- [30] E. M. Flanigen. Zeolites and molecular sieves: An historical perspective. *Stud. Surf. Sci. Catal.*, 137:11–15, 2001. doi:10.1016/S0167-2991(01)80243-3.
- [31] A. Gandhi and M. F. Hasan. Machine learning for the design and discovery of zeolites and porous crystalline materials. *Curr. Opin. Chem. Eng.*, 35:100739, 2022. doi:10.1016/j.coche.2021.100739.
- [32] A. C. Ghosh, A. Legrand, R. Rajapaksha, G. A. Craig, C. Sassoey, G. Balázs, D. Farusseng, S. Furukawa, J. Canivet, and F. M. Wisser. Rhodium-based metal–organic polyhedra assemblies for selective CO<sub>2</sub> photoreduction. *J. Am. Chem. Soc.*, 144(8):3626–3636, 2022. doi:10.1021/jacs.1c12631.
- [33] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, and Z. Wang. Hermit: an OWL 2 reasoner. *J. Autom. Reasoning.*, 53:245–269, 2014. doi:10.1007/s10817-014-9305-1.
- [34] A. Gogleva, D. Polychronopoulos, M. Pfeifer, V. Poroshin, M. Ughetto, M. J. Martin, H. Thorpe, A. Bornot, P. D. Smith, B. Sidders, J. R. Dry, M. Ahdesmäki, U. McDermott, E. Papa, and K. C. Bulusu. Knowledge graph-based recommendation framework identifies drivers of resistance in EGFR mutant non-small cell lung cancer. *Nat. Commun.*, 13:1667, 2022. doi:10.1038/s41467-022-29292-7.
- [35] S. R. Hall, F. H. Allen, and I. D. Brown. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallogr. Sect. A*, 47(6):655–685, Nov 1991. doi:10.1107/S010876739101067X.

- [36] A. Hedström. Ion exchange of ammonium in zeolites: a literature review. *J. Environ. Eng.*, 127(8):673–681, 2001. doi:10.1061/(ASCE)0733-9372(2001)127:8(673).
- [37] N. F. Himma, A. K. Wardani, N. Prasetya, P. T. Aryanti, and I. G. Wenten. Recent progress and challenges in membrane-based O<sub>2</sub>/N<sub>2</sub> separation. *Rev. Chem. Eng.*, 35(5):591–625, 2019. doi:10.1515/revce-2017-0094.
- [38] D. K. Kirtania. Chatgpt generated content and similarity index in chemistry. *J. Chem. Inf. Model.*, 2023. doi:10.1021/acs.jcim.3c01110.
- [39] L. Komissarov and T. Verstraelen. Zeo-1, a computational data set of zeolite structures. *Sci. Data*, 9(1):61, 2022. doi:10.1038/s41597-022-01160-5.
- [40] A. Kondinski and K. Y. Monakhov. Breaking the gordian knot in the structural chemistry of polyoxometalates: copper(II)–oxo/hydroxo clusters. *Chem. Eur. J.*, 23(33):7841–7852, 2017. doi:10.1002/chem.201605876.
- [41] A. Kondinski, A. Menon, D. Nurkowski, F. Farazi, S. Mosbach, J. Akroyd, and M. Kraft. Automated rational design of metal–organic polyhedra. *J. Am. Chem. Soc.*, 144(26):11713–11728, 2022. doi:10.1021/jacs.2c03402.
- [42] A. Kondinski, J. Bai, S. Mosbach, J. Akroyd, and M. Kraft. Knowledge engineering in chemistry: From expert systems to agents of creation. *Acc. Chem. Res.*, 56:128–139, 2023. doi:10.1021/acs.accounts.2c00617.
- [43] A. Kondinski, S. Mosbach, J. Akroyd, A. Breeson, Y. R. Tan, S. Rihm, J. Bai, and M. Kraft. Hacking decarbonization with a community-operated CreatorSpace. *Chem*, 2024. ISSN 2451-9294. doi:10.1016/j.chempr.2023.12.018.
- [44] N. B. Krdzavac, S. Mosbach, D. Nurkowski, P. Buerger, J. Akroyd, J. W. Martin, A. Menon, and M. Kraft. An ontology and semantic web service for quantum chemistry calculations. *J. Chem. Inf. Model.*, 59:3154–3165, 2019. doi:10.1021/acs.jcim.9b00227.
- [45] S. V. Krivovichev. Which inorganic structures are the most complex? *Angew. Chem. Int. Ed.*, 53(3):654–661, 2014. doi:10.1002/anie.201304374.
- [46] S. Kumar, R. Srivastava, and J. Koh. Utilization of zeolites as CO<sub>2</sub> capturing agents: Advances and future perspectives. *J. CO<sub>2</sub> Util.*, 41:101251, 2020. doi:10.1016/j.jcou.2020.101251.
- [47] Y. Li and J. Yu. New stories of zeolite structures: their descriptions, determinations, predictions, and evaluations. *Chem. Rev.*, 114(14):7268–7316, 2014. doi:10.1021/cr500010r.
- [48] Y. Li, L. Li, and J. Yu. Applications of zeolites in sustainable chemistry. *Chem*, 3(6):928–949, 2017. doi:10.1016/j.chempr.2017.10.009.
- [49] F. Liebau. Ordered microporous and mesoporous materials with inorganic hosts: definitions of terms, formula notation, and systematic classification. *Microporous Mesoporous Mater.*, 58(1):15–72, 2003. doi:10.1016/S1387-1811(02)00546-2.

- [50] M. Q. Lim, X. Wang, O. Inderwildi, and M. Kraft. *The World Avatar—A World Model for Facilitating Interoperability*, pages 39–53. Springer International Publishing, Cham, 2022. ISBN 978-3-030-86215-2. doi:10.1007/978-3-030-86215-2\_4.
- [51] A. Menon, L. Pascazio, D. Nurkowski, F. Farazi, S. Mosbach, J. Akroyd, and M. Kraft. Ontopesscan: An ontology for potential energy surface scans. *ACS Omega*, 8(2):2462–2475, 2023. doi:10.1021/acsomega.2c06948.
- [52] M. Moliner, Y. Román-Leshkov, and A. Corma. Machine learning applied to zeolite synthesis: the missing link for realizing high-throughput discovery. *Acc. Chem. Res.*, 52(10):2971–2980, 2019. doi:10.1021/acs.accounts.9b00399.
- [53] K. Möller and T. Bein. Mesoporosity—a new dimension for zeolites. *Chem, Soc. Rev.*, 42(9):3689–3707, 2013. doi:10.1039/C3CS35488A.
- [54] S. Mosbach, A. Menon, F. Farazi, N. Krdzavac, X. Zhou, J. Akroyd, and M. Kraft. Multiscale cross-domain thermochemical knowledge-graph. *J. Chem. Inf. Model.*, 60(12):6155–6166, 2020. doi:10.1021/acs.jcim.0c01145.
- [55] D.-K. Nguyen, V.-P. Dinh, H. Q. Nguyen, and N. T. Hung. Zeolite ZSM-5 synthesized from natural silica sources and its applications: a critical review. *J. Chem. Technol. Biotechnol.*, 98(6):1339–1355, 2023. doi:10.1002/jctb.7380.
- [56] S. P. Ong, S. Cholia, A. Jain, M. Brafman, D. Gunter, G. Ceder, and K. A. Persson. The materials application programming interface (api): A simple, flexible and efficient api for materials data based on representational state transfer (rest) principles. *Comput. Mater. Sci.*, 97:209–215, 2015. doi:10.1016/j.commatsci.2014.10.037.
- [57] OpenAI. Chatgpt (mar 14 version). <https://chat.openai.com/chat>, 2024. Large language model.
- [58] E. Pan, S. Kwon, Z. Jensen, M. Xie, R. Gómez-Bombarelli, M. Moliner, Y. Román-Leshkov, and E. Olivetti. Zeosyn: A comprehensive zeolite synthesis dataset enabling machine-learning rationalization of hydrothermal parameters. *ACS Cent. Sci.*, 2024. doi:10.1021/acscentsci.3c01615.
- [59] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu. Unifying large language models and knowledge graphs: A roadmap. *IEEE Trans. Knowl. Data Eng.*, pages 1–20, 2024. doi:10.1109/TKDE.2024.3352100.
- [60] L. Pascazio, S. Rihm, A. Naseri, S. Mosbach, J. Akroyd, and M. Kraft. Chemical species ontology for data integration and knowledge discovery. *J. Chem. Inf. Model.*, 63(21):6569–6586, 2023. doi:10.1021/acs.jcim.3c00820a.
- [61] L. Pascazio, D. Tran, S. Rihm, J. Bai, J. Akroyd, S. Mosbach, and M. Kraft. Question-answering system for combustion kinetics. Technical Report Technical Report 315, c4e-Preprint Series, Cambridge, 2023. URL <https://como.ceb.cam.ac.uk/preprints/315/>.

- [62] E. Pérez-Botella, S. Valencia, and F. Rey. Zeolites in adsorption processes: State of the art and future prospects. *Chem. Rev.*, 122(24):17647–17695, 2022. doi:10.1021/acs.chemrev.2c00140.
- [63] A. Primo and H. Garcia. Zeolites as catalysts in oil refining. *Chem. Soc. Rev.*, 43(22):7548–7561, 2014. doi:10.1039/C3CS60394F.
- [64] S. D. Rihm, J. Bai, A. Kondinski, S. Mosbach, J. Akroyd, and M. Kraft. Transforming research laboratories with connected digital twins. *Nexus*, 1(1):100004, 2024. doi:10.1016/j.ynxs.2024.100004.
- [65] H. Rijgersberg. OM - ontology of units of measure, 2023. URL <https://github.com/HajoRijgersberg/OM>. 2023.
- [66] H. Rijgersberg, M. van Assem, and J. Top. Ontology of units of measure and related concepts. *Semant. Web.*, 4:3–13, 2013. doi:10.3233/SW-2012-0069.
- [67] G. Saulius, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N. Serebryanaya, P. Moeck, R. Downs, and A. Bail. Crystallography open database—an open-access collection of crystal structures. *J. Appl. Crystallogr.*, 42(4):726–729, 2009. doi:10.1093/nar/gkr900.
- [68] D. Schwalbe-Koda and R. Gómez-Bombarelli. Generating, managing, and mining big data in zeolite simulations. *AI-Guided Design and Property Prediction for Zeolites and Nanoporous Materials*, pages 81–111, 2023. doi:10.1002/9781119819783.ch4.
- [69] J. Shin, D. Jo, and S. B. Hong. Rediscovery of the importance of inorganic synthesis parameters in the search for new zeolites. *Acc. Chem. Res.*, 52(5):1419–1427, 2019. doi:10.1021/acs.accounts.9b00073.
- [70] S. V. Smaalen. Incommensurate crystal structures. *Crystallogr. Rev.*, 4(2):79–202, 1995. doi:10.1080/08893119508039920.
- [71] B. Smit and T. L. Maesen. Molecular simulations of zeolites: adsorption, diffusion, and shape selectivity. *Chem. Rev.*, 108(10):4125–4184, 2008. doi:10.1021/cr8002642.
- [72] S. Staab and R. Studer. *Handbook on Ontologies*. Springer Verlag Berlin Heidelberg, 2004. doi:10.1007/978-3-540-24750-0.
- [73] K. B. Tankersley, N. P. Dunning, C. Carr, D. L. Lentz, and V. L. Scarborough. Zeolite water purification at tikal, an ancient maya city in guatemala. *Sci. Rep.*, 10:18021, 2020. doi:10.1038/s41598-020-75023-7.
- [74] G. Tecuci, D. Marcu, M. Boicu, and D. A. Schum. *Knowledge engineering: Building cognitive assistants for evidence-based reasoning*. Cambridge University Press, 2016.

- [75] D. Tran, L. Pascazio, J. Akroyd, S. Mosbach, and M. Kraft. Leveraging text-to-text pretrained language models for question answering in chemistry. *ACS Omega*, 9(9): 13883–13896, 2024. doi:10.1021/acsomega.3c08842.
- [76] M. Treacy and J. Higgins, editors. *Collection of Simulated XRD Powder Patterns for Zeolites (Fifth Edition)*. Elsevier Science B.V., Amsterdam, fifth edition edition, 2007. ISBN 978-0-444-53067-7. doi:10.1016/B978-044453067-7/50472-X.
- [77] S. van Smaalen. Symmetry of composite crystals. *Phys. Rev. B*, 43:11330–11341, May 1991. doi:10.1103/PhysRevB.43.11330.
- [78] V. Van Speybroeck, K. Hemelsoet, L. Joos, M. Waroquier, R. G. Bell, and C. R. A. Catlow. Advances in theory and their application within the field of zeolite chemistry. *Chem. Soc. Rev.*, 44(20):7044–7111, 2015. doi:10.1039/C5CS00029G.
- [79] H. Wang, Y. Xie, D. Li, H. Deng, Y. Zhao, M. Xin, and J. Lin. Rapid identification of x-ray diffraction patterns based on very limited data by interpretable convolutional neural networks. *J. Chem. Inf. Model.*, 60(4):2004–2011, 2020. doi:10.1021/acs.jcim.0c00020.
- [80] Y. Wu and B. M. Weckhuysen. Separation and purification of hydrocarbons with porous materials. *Angew. Chem. Int. Ed.*, 60(35):18930–18949, 2021. doi:10.1002/anie.202104318.
- [81] H. Xu and P. Wu. New progress in zeolite synthesis and catalysis. *Natl. Sci. Rev.*, 9(9):nwac045, 2022. doi:10.1093/nsr/nwac045.
- [82] O. M. Yaghi. Reticular chemistry in all dimensions. *ACS Cent. Sci.*, 5(8):1295–1300, 2019. doi:10.1021/acscentsci.9b00750.
- [83] S. Yang, M. Lach-Hab, I. I. Vaisman, E. Blaisten-Barojas, X. Li, and V. L. Karen. Framework-type determination for zeolite structures in the inorganic crystal structure database. *J. Phys. Chem. Ref. Data.*, 39(3), 2010. doi:10.1063/1.3432459.
- [84] B. Yue, S. Liu, Y. Chai, G. Wu, N. Guan, and L. Li. Zeolites for separation: Fundamental and application. *J. Energy Chem.*, 71:288–303, 2022. doi:10.1016/j.jechem.2022.03.035.
- [85] C. Zheng, Y. Li, and J. Yu. Database of open-framework aluminophosphate structures. *Sci. Data*, 7(1):107, 2020. doi:10.1038/s41597-020-0452-4.
- [86] N. Zheng, X. Bu, B. Wang, and P. Feng. Microporous and photoluminescent chalcogenide zeolite analogs. *Science*, 298(5602):2366–2369, 2002. doi:10.1126/science.1078663.
- [87] X. Zhou, D. Nurkowski, S. Mosbach, J. Akroyd, and M. Kraft. Question answering system for chemistry. *J. Chem. Inf. Model.*, 61(8):3868–3880, 2021. doi:10.1021/acs.jcim.1c00275.

- [88] X. Zhou, D. Nurkowski, A. Menon, J. Akroyd, S. Mosbach, and M. Kraft. Question answering system for chemistry - a semantic agent extension. *Digital Chem. Eng.*, 3:100032, 2022. doi:10.1016/j.dche.2022.100032.
- [89] N. E. R. Zimmermann and M. Haranczyk. History and utility of zeolite framework-type discovery from a data-science perspective. *Cryst. Growth Des.*, 16(6):3043–3048, 2016. doi:10.1021/acs.cgd.6b00272.



# Supporting Information

## SI.1 Ontology Design and Data Details

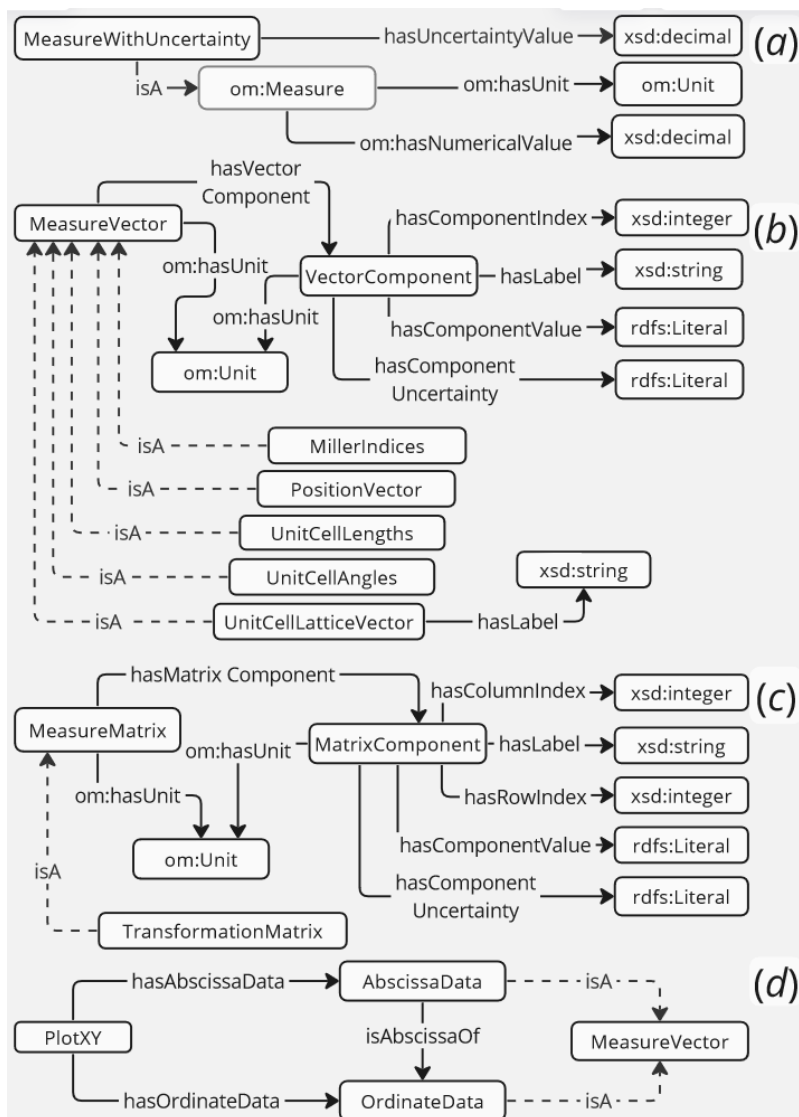
The measured data in the Crystal Information File (CIF) typically have the uncertainty of measurement, which is not yet supported in the most recent version of the Units Of Measurement Ontology (OM2) [65, 66]. We introduce a `ocr:MeasureWithUncertainty` class Fig. S1.a which is an extension of the `om:Measure` class with additional property `hasUncertaintyValue`. This data property shares the same unit of measurement with the `hasNumericalValue` property.

Also, the CIF format contains a variety of array-like data types, and it is convenient to store such data as components of a single structure. Although not all of these sets are vectors in strict mathematical meaning, we name the new class as `ocr:MeasureVector`, to highlight that this data structure is a set of `om:Measure` values with value, units of measurement. This data structure is flexible in terms of the number of components, labeling of the components, individual units of measurement for the components, and may contain uncertainties. Depending on the implementation the components can be accessed by the index and/or by the label of the component. Same applies to the components of `ocr:MeasureMatrix`. Large data sets, like spectrum plots, can be represented by two arrays of points for the X and Y axes, measured in their respective units. The data arrays are stored as `MeasureVector`, and may contain uncertainties to represent the plot with error bars.

## Data details

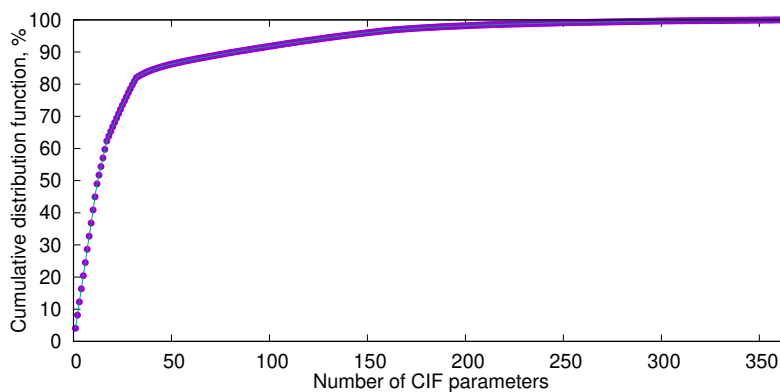
CIF files cover the majority of the crystal properties and details on the measurement procedure. The total count of keywords reaches over 3000 in the recent standard [10], plus it allows non-standard keywords defined by software and hardware developers. Within this extensive list, certain keywords emerge as more prevalent than others. We studied the occurrence of the keywords in the CIF files in the COD database [67]. As COD contains CIF files from a variety of sources, we assume that this is a good representation of CIF files in general. Figure S2 shows the cumulative distribution function for the top 350 keywords. Our analysis shows that the top 50 keywords cover over 80% of the information contained in the CIF files. These top keywords include the crystal unit cell data, the atomic structure, and the citation. The OntoCrystal ontology mainly focuses on these top keywords as the most commonly used. Other keywords may be added in the later extensions of the ontology.

Zeolite as a crystalline material contains crystal information from CIF is a subset. In addition it includes information on the porosity, chemical compounds, framework structure, etc. Therefore the OntoZeolite ontology covers the topological properties, tiling structure, the framework. The Knowledge Graph currently includes XRD spectra for 175 zeolite frameworks, based on [76]. The framework codes included in the KG are listed below:



**Figure S1:** Ontology structure of data type defined in OntoCrystal. (a) *MeasureWithUncertainty* is an extension of the *om:Measure*, (b) *MeasureVector* is a data structure made of several *om:Measure* values with uncertainties, (c) *MeasureMatrix* is a 2D array of the *om:Measure* with uncertainties, (d) *PlotXY* is a 2D plot, where both abscissa and ordinate are *MeasureVector* data. The components of the vectors and matrices can be accessed by an index or label.

ABW, ACO, AEI, AEL, AEN, AET, AFG, AFI, AFN, AFO, AFR, AFS, AFT, AFX, AFY, AHT, ANA, APC, APD, AST, ASV, ATN, ATO, ATS, ATT, ATV, AWO, AWW, BCT, BEC, BIK, BOG, BPH, BRE, CAN, CAS, CDO, CFI, CGF, CGS, CHA, CHI, CLO, CON, CZP, DAC, DDR, DFO, DFT, DOH, DON, EAB, EDI, EMT, EON, EPI, ERI, ESV, ETR, EUO, EZT, FAR, FAU, FER, FRA, GIS, GIU, GME, GON, GOO, HEU, IFR, IHW, IMF, ISV, ITE, ITH, ITW, IWR, IWV, IWW, JBW, KFI, LAU, LEV, LIO, LIT, LOS, LOV, LTA, LTL, LTN, MAR, MAZ, MEI, MEL, MEP, MER, MFI, MFS, MON, MOR, MOZ,

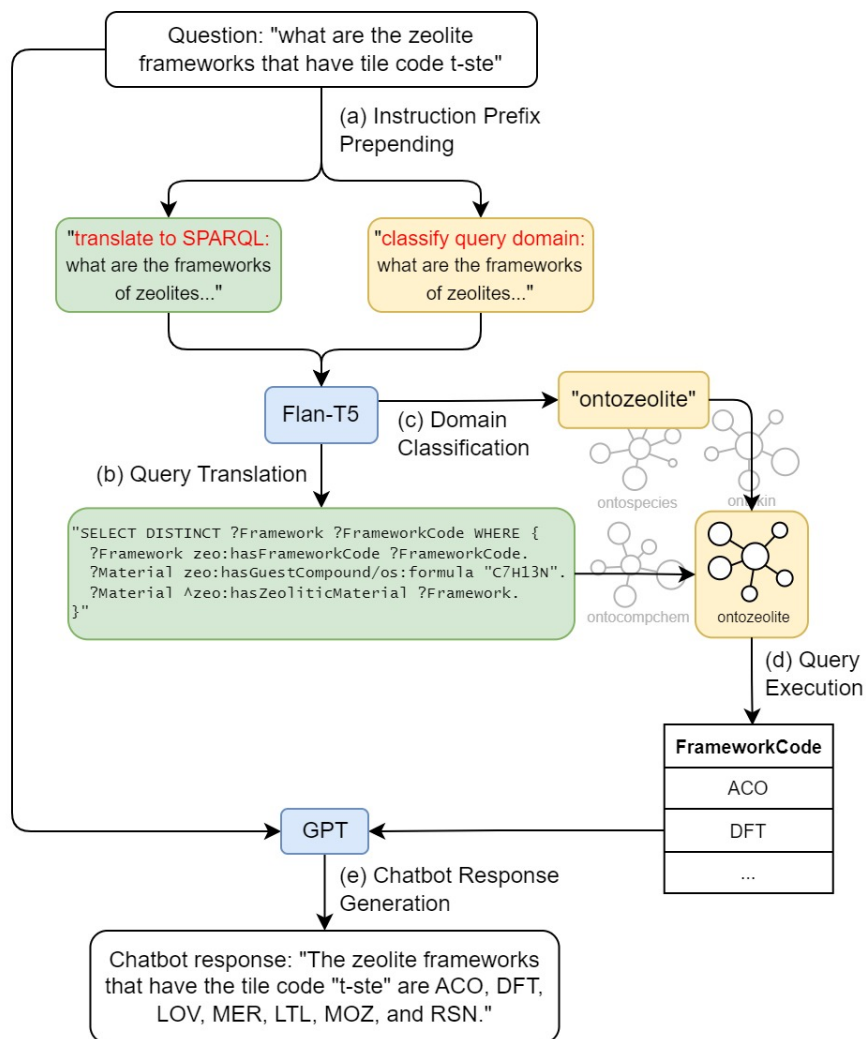


**Figure S2:** *Analysis of Crystallographic Information Files (CIF). The cumulative distribution function for the most common keywords in the COD database. Top 50 keywords cover over 80% of the data contained in the CIF files.*

MSE, MSO, MTF, MTN, MTT, MTW, MWW, NAB, NAT, NES, NON, NPO, NSI, OBW, OFF, OSI, OSO, OWE, PAR, PAU, PHI, PON, RHO, RON, RRO, RSN, RTE, RTH, RUT, RWR, RWY, SAO, SAS, SAT, SAV, SBE, SBS, SBT, SFE, SFF, SFG, SFH, SFN, SFO, SGT, SIV, SOD, SOS, SSY, STF, STI, STT, SZR, TER, THO, TOL, TON, TSC, TUN, UEI, UFI, UOZ, USI, UTL, VET, VFI, VNI, VSV, WEI, WEN, YUG, ZON.

All codes related to the conversion of files are accessible on The World Avatar Git repository (<https://github.com/cambridge-cares/TheWorldAvatar>). Data files suitable for the regeneration of the knowledge graph are made accessible through the University of Cambridge Data Repository.

## SI.2 NLP Workflow, Questions and Responses



**Figure S3:** Workflow demonstrating the translation of a natural language question about zeolite frameworks into a SPARQL query, including query execution and chatbot response generation.

## Marie Responses

### What is the unit cell information of zeolite framework ...

**User:** What is the unit cell information of zeolitic framework ABW?

**Marie's response:**

The unit cell information of zeolitic framework ABW includes the following details:

- Lengths:
  - $a$ : 9.873 Å
  - $b$ : 5.254 Å
  - $c$ : 8.77 Å
- Angles:
  - alpha: 90.0°
  - beta: 90.0°
  - gamma: 90.0°
- Volume: 454.9239 cubic Ångstroms
- Lattice system: Orthorhombic
- Space group number (ITCr): 74

Correct Response

This response is correct. For comparison see [IZA - Framework Type ABW](#)

**User:** What is the unit cell information of zeolite AHT?

**Marie response:**

The unit cell information of zeolite AHT is as follows:

- Lengths:  $a = 15.794$  Å,  $b = 9.206$  Å,  $c = 8.589$  Å
- Angles:  $\alpha = 90.0^\circ$ ,  $\beta = 90.0^\circ$ ,  $\gamma = 90.0^\circ$
- Volume: 1248.8369 cubic Å
- Lattice System: Orthorhombic
- Space Group Number (ITCr): 63

Correct Response

This response is correct. For comparison see [IZA - Framework Type AHT](#)

**User:** What is the unit cell information of zeolite LAU?

**Marie response:**

The unit cell information of zeolite LAU is as follows:

- Lengths: The lengths of the unit cell are  $a = 14.587 \text{ \AA}$ ,  $b = 12.877 \text{ \AA}$ ,  $c = 7.613 \text{ \AA}$ .
- Angles: The angles of the unit cell are  $\alpha = 90.0^\circ$ ,  $\beta = 111.159^\circ$ ,  $\gamma = 90.0^\circ$ .
- Volume: The volume of the unit cell is 1333.5942 cubic  $\text{\AA}$ .
- Lattice System: The zeolite LAU belongs to the monoclinic lattice system.
- Space Group Number: The space group number (according to the International Tables of Crystallography) for zeolite LAU is 12.

Correct Response

This response is correct. For comparison see [IZA - Framework Type AHT](#)

## Show me zeolite frameworks incorporating ...

**User:** Show me zeolite frameworks incorporating pyridine.

**Marie's response:** Zeolite frameworks incorporating pyridine include RFE, FER, and CHA.

Correct Response

This response is correct. For comparison see. Figure S4 documents the IZA response for comparison.

Search Results - Composition of Materials

Search criteria: OSDA = C<sub>5</sub>H<sub>5</sub>N = pyridine  
Materials found: 1  
Output sorted by: Framework Type Code in ascending order

FTC	Material	Si/Al	Short Fw Formula	Stoichiometry (if available)
RFE	RUB-58, as-made		[Si - O]	[Pyr <sub>1.33</sub> EDA <sub>0.67</sub> ][Si <sub>48</sub> O <sub>96</sub> ]-RFE, Pyr = C <sub>5</sub> H <sub>5</sub> N = pyridine, EDA = C <sub>2</sub> H <sub>8</sub> N <sub>2</sub> = ethylenediamine

Copyright © 2020 Structure Commission of the International Zeolite Association (IZA-SC)

**Figure S4:** IZA Search for pyridine as guest molecule. Query from 14 April 2024.

## Find zeolites with framework element components ...

**User:** Find zeolites with framework element components Ga and Se only.

### Marie's response:

I found two zeolites with framework element components Ga and Se only:

1. Chemical Formula:  $[Ga_4Se_8]$
2. Chemical Formula:  $[(CID85631)][Ga_4Se_8]$

These zeolites have Ga and Se as their framework element components.

The screenshot shows a search interface with a search bar containing the query "Find zeolites with framework element components Ga and Se only". Below the search bar, there are two buttons: "Click to view the predicted SPARQL query" and "Click to view the post-processed SPARQL query". A "Hide IRIs" button is also present. The search results are displayed in a table with columns for "#", "Material", and "ChemicalFormula".

#	Material	ChemicalFormula
1	<a href="http://www.theworldavatar.com/kg/ontozeolite/ZeoIiticMaterial_4d09f275-af0e-480b-8f87-04780571bc46">http://www.theworldavatar.com/kg/ontozeolite/ZeoIiticMaterial_4d09f275-af0e-480b-8f87-04780571bc46</a>	[Ga4Se8]
2	<a href="http://www.theworldavatar.com/kg/ontozeolite/ZeoIiticMaterial_50294a38-280a-4139-abee-37e14f0bc824">http://www.theworldavatar.com/kg/ontozeolite/ZeoIiticMaterial_50294a38-280a-4139-abee-37e14f0bc824</a>	[(CID85631)][Ga4Se8]

Showing 1 to 2 of 2 entries

IRI details for the first result:

**IRI:** [http://www.theworldavatar.com/kg/ontozeolite/ZeoIiticMaterial\\_4d09f275-af0e-480b-8f87-04780571bc46](http://www.theworldavatar.com/kg/ontozeolite/ZeoIiticMaterial_4d09f275-af0e-480b-8f87-04780571bc46)  
**doi:** 10.1126/science.1078663  
**Zeolite Framework:** RWY  
**Chemical Formula:** [Ga4Se8]

Crystallographic parameters for the first result:

P 1 [P 1] #1  
a=21.893Å  
b=21.893Å  
c=21.893Å  
α=90.000°  
β=90.000°  
γ=90.000°

**Figure S5:** IZA Search for pyridine as guest molecule. Query from 14 April 2024.

### Correct Response

This answer and reference are correct, and they link to an influential work published in DOI: [10.1126/science.1078663](https://doi.org/10.1126/science.1078663).

## ChatGPT Responses

### What is the unit cell information of zeolite framework ...

**User:** What is the unit cell information of zeolite ABW?

**ChatGPT:**

Zeolite ABW, one of the many types of zeolite frameworks, has a unique unit cell structure characterized by specific lattice parameters and symmetry. Here are the details of the zeolite ABW unit cell:

- Crystal System: Hexagonal
- Space Group: P6<sub>3</sub>/mmc
- Lattice Parameters:
  - $a =$  about 10.4 Å
  - $c =$  about 6.4 Å

This framework consists of a three-dimensional network of aluminum, silicon, and oxygen that forms channels and cavities. These properties make it useful for various applications including ion exchange, adsorption, and catalysis.

#### Incorrect Response

The chatbot hallucinates. A shared response can be found on the following link: <https://chat.openai.com/share/64e1673a-ef9e-484a-882e-f29939c7f058> For comparison see [IZA - Framework Type ABW](#)



**User:** What is the unit cell information of zeolite framework AHT?

**ChatGPT:** The zeolite framework type designated as "AHT" refers to a specific structure within the family of zeolite materials. Zeolites are microporous, aluminosilicate minerals commonly used as commercial adsorbents and catalysts. The framework type of a zeolite is determined by its unique three-dimensional arrangement of silica and alumina tetrahedra.

The "AHT" framework type is less commonly discussed compared to more well-known zeolites like Type A (LTA), Type X, and Type Y. To provide accurate unit cell information for the AHT zeolite framework, including its dimensions, symmetry, and typical chemical composition, I'll need to check some detailed resources. Let me do that right now.

The zeolite framework type AHT has specific unit cell information detailed in the databases of the International Zeolite Association. The AHT structure has a monoclinic unit cell with the following dimensions and angles:  $a = 9.4860 \text{ \AA}$ ,  $b = 9.9140 \text{ \AA}$ ,  $c = 8.1260 \text{ \AA}$ ,  $\alpha = 90.000^\circ$ ,  $\beta = 90.000^\circ$ ,  $\gamma = 121.490^\circ$ . This framework is characterized by its framework density and specific geometric configuration suitable for various applications.

Another specification for the AHT framework is its orthorhombic cell parameters under a different configuration or material composition:  $a = 15.7940 \text{ \AA}$ ,  $b = 9.2060 \text{ \AA}$ ,  $c = 8.5890 \text{ \AA}$ , with all angles ( $\alpha, \beta, \gamma$ ) at  $90.000^\circ$ , giving a volume approximately  $1248.8 \text{ \AA}^3$ . This demonstrates the versatility and structural variations within the AHT type frameworks depending on the specific synthesis and material composition.

These details highlight the complex nature of zeolite framework structures, which are crucial for their function in catalysis and adsorption processes due to the precise arrangement of pores and channels within the material.

For more detailed information and visual representations of the AHT framework, you can visit the IZA Structure Commission's database: [Zeolite AHT Framework](#).

#### Partially Correct Response

The chatbot originally hallucinates, but the second part of the response provides correct information on the framework parameters. The chatbot provides the correct link to the IZA webpage. A shared response can be found on the following link: <https://chat.openai.com/share/443d6bcb-228e-40e5-b68c-a1b8a9643e36>

**User:** What is the unit cell information of zeolite LAU?

**ChatGPT:** Zeolite LAU, also known as Linde Type A or LTA, is a synthetic zeolite with a well-defined structure and specific properties that make it suitable for various industrial applications, especially in adsorption and ion exchange processes. Here's the unit cell information for zeolite LAU (LTA):

- **Chemical Formula:** The chemical formula of zeolite LAU is typically given as  $\text{Na}_{12}[\text{Al}_{12}\text{Si}_{12}\text{O}_{48}] \cdot 27\text{H}_2\text{O}$ .
- **Crystal System:** Zeolite LAU crystallizes in the cubic crystal system.
- **Unit Cell Dimensions:** The unit cell parameter  $a$  is approximately 24.60 Å.
- **Framework Type:** The framework type of zeolite LAU is designated as LTA. It features a three-dimensional network of  $\text{AlO}_4$  and  $\text{SiO}_4$  tetrahedra linked together to form a characteristic sodalite cage structure. This arrangement creates a system of pores and channels, ideal for molecular sieving.
- **Space Group:** The space group for zeolite LAU is  $\bar{3}m$ , indicating a highly symmetric structure.
- **Pore Dimensions:** It has uniform pore openings of about 4 Å, making it particularly useful for size-selective adsorption of small molecules.

Zeolite LAU is utilized in processes like water softening and the separation of small molecules due to its ability to selectively adsorb molecules that fit within its pores, which is largely determined by the aforementioned structural characteristics.

#### Incorrect Response

The chatbot clearly mixes framework-type LAU with framework-type LTA. A shared response can be found on the following link: <https://chat.openai.com/share/25dcedbd-f2ba-40de-a619-83942e29d82e>

## Show me zeolite frameworks incorporating ...

**User:** Show me zeolite frameworks incorporating pyridine.

**ChatGPT:** Zeolites with incorporated pyridine represent a specialized topic within materials science, focusing on modifying the properties of zeolite frameworks through the introduction of organic molecules like pyridine. Pyridine is a basic, nitrogen-containing heterocycle that can interact with the acidic sites of zeolites, potentially affecting the catalytic and adsorptive properties of the zeolite.

Incorporating pyridine into zeolite frameworks can be done in several ways:

**Synthesis Modification:** Pyridine can be added during the synthesis of the zeolite to direct the assembly of the zeolite framework, acting as a structure-directing agent (SDA). This incorporation can influence the pore size, topology, and framework stability.

**Post-synthesis Modification:** Pyridine can be introduced into pre-formed zeolites through ion-exchange processes, where pyridine molecules replace other cations in the zeolite's structure. This modification can enhance the selectivity and efficiency of zeolites in various reactions.

**Functionalization and Adsorption:** Pyridine can also be adsorbed onto the surface of zeolites, modifying the surface characteristics and interacting with reactant molecules in catalytic processes.

These modifications aim to enhance the utility of zeolites in applications such as catalysis (e.g., cracking, isomerization, and fine chemical synthesis), gas separation, and pollution control.

For visualization or specific structures of these frameworks, diagrams or modeling software like those used in materials science or chemistry would be helpful to see how pyridine is arranged within the zeolite matrix. If you're looking for more detailed explanations or visual materials, specific research articles or textbooks on zeolite chemistry would be ideal sources.

### Incorrect/Partially Correct Response

The chatbot hallucinates and does not answer the question asked. A shared response can be found on the following link: <https://chat.openai.com/share/6f312790-5401-4568-a7c3-b66c2c2f8a6f> Rerunning the prompt links provides to NMR studies linking pyridine to HZSM-5 (DOI: 10.1039/D1CP01515J) <https://chat.openai.com/share/04686096-ee95-440e-ad15-5c3a43b3413d>

**User:** What is the tile information of zeolite framework SOD?

**ChatGPT:** The "SOD" in zeolite framework refers to sodalite, a type of zeolite mineral characterized by its cubic framework. The "SOD" topology in zeolites has a three-dimensional structure composed of six-membered rings of SiO<sub>4</sub> and AlO<sub>4</sub> tetrahedra. This forms a distinctive cubic lattice.

The tile information, which describes the types of secondary building units (SBUs) used to construct the framework, includes the following for the SOD framework:

- Primary Building Unit (PBU): The SOD framework is typically built from a primary building unit consisting of  $\beta$ -cages, which are cubic units made up of 24 tetrahedral units (T-atoms).
- Secondary Building Unit (SBU): The SOD framework does not have traditional larger SBUs like some other zeolites but is primarily constructed from the interconnected  $\beta$ -cages.

The sodalite cage itself is a truncated octahedron and is a key characteristic unit in many zeolites, not just those with the SOD framework. These cages are connected in such a way that they form a dense, yet porous structure, ideal for various applications including ion-exchange, gas separation, and catalysis.

#### Correct/partially Response

The sodalite framework contains only a single tiling unit also called sodalite or  $\beta$ -cage. This cage is well known to be a truncated octahedron. The response is available through the following link: <https://chat.openai.com/share/42d3c2d9-85be-4020-ba42-ef35dc8ccd69>

**User:** What is the tile information of zeolite framework CGF?

**ChatGPT:** The zeolite framework type CGF, abbreviated from "Chabazite Germanium Framework," is a hypothetical ordered zeolite framework constructed through simulations by substituting the silicon atoms in the CHA (chabazite) framework type with germanium. In the context of zeolite frameworks, "tiles" often refer to the basic building blocks that make up the framework structure, which could be rings or secondary building units (SBUs) such as polyhedral clusters.

The tile information of the CGF framework can be described as follows:

- Framework Code: CGF
- Topology: This framework is related to the chabazite (CHA) type but with germanium substitution.
- Tile Composition: The tiles in CGF would likely include 4-, 6-, and 8-membered rings, similar to those found in the CHA framework.
- Dimensionality: It is a three-dimensional framework.
- Pore System: The framework contains pores defined by the arrangement and connectivity of its tiles.

These hypothetical zeolites like CGF often explore substitutions of atoms like silicon with other tetrahedral atoms (like germanium) to predict new material properties or understand the structural impacts of such substitutions. The tile-based description is crucial for understanding the pore structure and potential applications in catalysis, gas separation, or ion exchange.

#### Incorrect/Partially Correct Response

This answer is a hallucination. The abbreviation CGF derives from **C**obalt-**G**allium-**P**hosphate-**F**ive. Consequently information related to tiling is incorrect. The response can be accessed via the link: <https://chat.openai.com/share/87dc6e33-fe50-4341-9b77-5a7f55df1b3e>

## Find zeolites with framework element components ...

**User:** Find zeolites with framework element components Ga and Se only.

**ChatGPT:** Zeolites are microporous, aluminosilicate minerals commonly used in industrial applications such as catalysis, ion exchange, and gas separation. The structure of a zeolite is characterized by its framework, which consists of tetrahedrally coordinated atoms linked by oxygen atoms.

To find a zeolite that specifically uses gallium (Ga) and selenium (Se) as the framework components is quite unique. Typically, zeolites do not include selenium in their framework because it is not commonly tetrahedrally coordinated in minerals. Selenium usually prefers other types of coordination geometries due to its larger size and different electronic properties compared to silicon or aluminum, which are more commonly found in zeolite frameworks.

Gallium, however, can substitute for aluminum in some zeolite frameworks due to its similar size and charge. It is known to be used in some special synthetic zeolites for specific catalytic or adsorption properties.

Given the unconventional nature of using selenium in a zeolite framework, it is unlikely that such a zeolite with only Ga and Se as framework elements exists based on common knowledge and typical properties of zeolite structures. If such materials were to be synthesized or discovered, they would likely be highly novel and possibly reported in specialized scientific research focused on advanced materials or novel crystallography.

For accurate and current information, searching through scientific literature databases such as SciFinder or Web of Science or reading recent publications in materials science journals would be necessary. These platforms would provide the most up-to-date research on the synthesis and properties of such unique zeolitic materials, if they exist.

### Lack of information response

This answer suggests that the chatbot does not have information that such zeolite structural composition exists. The response can be accessed via the link: <https://chat.openai.com/share/33187465-1dc6-4629-91ed-28d628cded70>

## Gemini Advanced Responses

### What is the unit cell information of zeolitic framework ...

**User:** What is the unit cell information of zeolitic framework ABW?

**Gemini Advanced:** Here's the unit cell information of the zeolitic framework ABW:

**Framework Type:** ABW

**Crystal System:** Orthorhombic

**Space Group:** Pna21 (No. 33)

**Typical Unit Cell Parameters:**

- $a = 10.3 \text{ \AA}$
- $b = 8.2 \text{ \AA}$
- $c = 5.0 \text{ \AA}$

**Important Notes:**

- **Composition:** The ABW framework is typically formed with aluminosilicate and gallosilicate compositions. The specific unit cell parameters can slightly vary depending on the exact Si/Al or Si/Ga ratio.
- **Structure:** The ABW framework is known for its small pore system, limiting the size of molecules that can enter it. The framework structure features a combination of 4-rings, 6-rings, and 8-rings.

**Database References:** You can find detailed information about the ABW framework topology, including unit cell data, on the International Zeolite Association (IZA) Structure Commission website: <http://www.iza-structure.org/databases/>

**Where to Find More Information:**

- **Atlas of Zeolite Framework Types:** This comprehensive resource provides detailed structural descriptions and illustrations of zeolite frameworks, including ABW. ([https://books.google.co.uk/books?id=c5\\_JGBc7VZgC&pg=PA21&lpg=PA21&dq=unit+cell+information+of+zeolitic+framework+ABW&source=bl&ots=xUsyysbLI-&sig=ACfU3U0jG870lC3zbNnVvmYYzsvWnknB0g&hl=en](https://books.google.co.uk/books?id=c5_JGBc7VZgC&pg=PA21&lpg=PA21&dq=unit+cell+information+of+zeolitic+framework+ABW&source=bl&ots=xUsyysbLI-&sig=ACfU3U0jG870lC3zbNnVvmYYzsvWnknB0g&hl=en))
- **Research Publications:** Check scientific journals and databases for publications on zeolite synthesis and characterization for more specific experimental unit cell parameters and structural analyses of ABW zeolites.

Let me know if you'd like more information on the ABW framework!

#### Lack of information response

The chatbot indeed provides link to the IZA database, however, the provided data for ABW are made up and do not correspond to what is being reported there. The response can be accessed via the link: <https://g.co/gemini/share/aa8a2031e241>

**User:** What is the unit cell information of zeolite AHT?

**Gemini Advanced:** Here's a breakdown of the unit cell information for zeolite AHT:

**Structure and Framework Type**

- **Framework Type Code:** AHT (assigned by the International Zeolite Association's Structure Commission)
- **Construction:** AHT is built from unique secondary building units (SBUs) called the 10-ring tube. These tubes connect to each other to form the overall framework.
- **Topology:** The AHT framework has a complex topology with interconnected channels and cages.

**Unit Cell Parameters**

- **Crystal System:** Tetragonal
- **Space Group:** I4/mmm (No. 139)
- **Lattice Parameters:**
  - $a \approx 18.9 \text{ \AA}$
  - $c \approx 8.4 \text{ \AA}$
- **Unit Cell Volume:** Approximately  $3000 \text{ \AA}^3$
- **Zeolite formula per unit cell:** The specific formula depends on the cations present, but a general form is  $|M_{12}(H_2O)_{27}|[Al_{12}Si_{36}O_{96}]$ , where  $M$  represents exchangeable cations.

**Key Characteristics**

- **Pore System:** Zeolites AHT has a pore system formed by interconnected channels with 10-membered ring openings.
- **Channel Dimensions:** The approximate pore size is around  $4.4 \times 7.2 \text{ \AA}$ .

**Important Notes:**

- Unit cell parameters can slightly vary based on variations in chemical composition or synthesis conditions. The values provided above are representative.
- Zeolites are complex materials, and detailed structural information is best obtained from crystallographic databases and research papers.

**Where to Find More Information**

- International Zeolite Association Structure Database: <http://www.iza-structure.org/databases/>
- Research publications: Search for research articles specifically focusing on the structure and properties of zeolite AHT.

Let me know if you'd like any specific details or have further questions!

**Lack of information response**

The chatbot indeed provides link to the IZA database, however, the provided data for AHT are made up and do not correspond to what is being reported there. The response can be accessed via the link: <https://g.co/gemini/share/b1f7ff2d5365>



**User:** What is the unit cell information of zeolite LAU?

**Gemini Advanced:** Zeolite LAU (Laumontite) has the following unit cell information:

**Crystal System:** Monoclinic **Lattice Parameters:**

- $a \approx 13.7 \text{ \AA}$
- $b \approx 13.1 \text{ \AA}$
- $c \approx 7.5 \text{ \AA}$
- $\alpha = 90^\circ$
- $\beta \approx 112^\circ$
- $\gamma = 90^\circ$

**Space group:** C2/m

**Important Notes:**

- Zeolite structures, including laumontite, can be somewhat flexible. Variations in hydration state or chemical substitutions can slightly alter unit cell dimensions.
- Laumontite is known to transition to a lower symmetry form upon dehydration.

**Key Structural Features of the LAU Framework**

- **Building Units:** The LAU framework consists of 4-rings of tetrahedra (where each tetrahedron commonly represents a  $\text{SiO}_4$  or  $\text{AlO}_4$  unit). These combine to form larger structural units such as double 6-rings.
- **Channels:** The LAU framework has a system of one-dimensional channels running parallel to the c-axis.
- **Cages:** The structure contains 'lau' cages that are accessible through the channels.

**Where to Find More Information**

- International Zeolite Association Structure Database: This is the authoritative source for zeolite framework information (<http://www.iza-structure.org/databases/>)
- Scientific Publications: Research articles on the synthesis, properties, and modification of laumontite often detail its crystallographic information.

Let me know if you'd like more details on the LAU structure or zeolites in general!

#### Incorrect Data

The chatbot indeed provides link to the IZA database, however, the provided data for LAU are made up and do not correspond to what is being reported there. The response can be accessed via the link: <https://g.co/gemini/share/f49751046133>

## Show me zeolite frameworks incorporating ...

**User:** Show me zeolite frameworks incorporating pyridine.

### **Gemini Advanced:**

Unfortunately, it's difficult to find examples of zeolite frameworks that directly incorporate pyridine within their structure. Here's why and some alternative ways pyridine is used with zeolites:

#### **Challenges of Direct Pyridine Incorporation**

- **Size and Shape Mismatch:** Pyridine is a relatively bulky molecule compared to the typical building blocks of zeolite frameworks (silica and alumina tetrahedra). Incorporating it directly would disrupt the zeolite's regular structure.
- **Stability:** The nitrogen atom in pyridine can act as a Lewis base. This basicity could make the zeolite unstable in acidic or reactive environments, as the nitrogen could interact with other molecules.

#### **How Pyridine is Used with Zeolites**

- **Templating Agent:** Pyridine-like molecules can be used during zeolite synthesis to direct the formation of specific pore structures. The pyridine-like molecule is later removed, leaving behind a zeolite with a tailored structure.
- **Post-Synthesis Modification:** Pyridine can be attached to existing zeolites through chemical reactions. This creates functionalized zeolites with sites for:
  - **Catalysis:** Pyridine's basic nitrogen can participate in catalytic reactions.
  - **Adsorption:** Pyridine-functionalized zeolites can selectively adsorb molecules that interact with the nitrogen.
- **Guest Molecule:** Pyridine can be trapped within zeolite pores without being chemically bound to the framework. This is useful for:
  - **Catalysis:** Confining pyridine in the zeolite can influence its reactivity in catalytic processes.
  - **Molecular Sensing:** The interaction of pyridine with the zeolite framework can be used for sensing specific molecules.

**Note:** While it's less common to directly incorporate pyridine into the framework, there are zeolite-like materials called Metal-Organic Frameworks (MOFs) that are much more flexible in terms of the organic molecules they can incorporate. You might find examples of MOFs with integrated pyridine components.

Irrelevant response

The response can be accessed via the link: <https://g.co/gemini/share/52dad5f8579b>

## Find zeolites with framework element components ...

**User:** Find zeolites with framework element components Ga and Se only.

### **Gemini Advanced:**

Unfortunately, it's highly unlikely that zeolites with a framework composed solely of gallium (Ga) and selenium (Se) exist. Here's why:

#### **Zeolites Fundamentals**

- **Definition:** Zeolites are crystalline microporous aluminosilicates. This means their framework primarily consists of aluminum (Al), silicon (Si), and oxygen (O).
- **Tetrahedral Building Blocks:** The fundamental building units in zeolites are tetrahedra where a silicon or aluminum atom sits at the center, surrounded by four oxygen atoms. These tetrahedra link together to form the intricate zeolite frameworks.
- **Charge Balancing:** Aluminum has a +3 charge, while silicon has a +4 charge. This charge imbalance necessitates the presence of cations (often sodium, potassium, etc.) within the zeolite structure to maintain neutrality.

#### **Why Ga and Se Are Unlikely**

- **Charge Considerations:** Gallium (Ga), like aluminum, has a +3 charge. Selenium (Se) can have varying oxidation states (-2, +4, +6). It's difficult to envision a stable zeolite framework where Ga and Se are the only elements present while maintaining charge neutrality.
- **Coordination Preference:** Gallium tends to favor tetrahedral coordination (like aluminum and silicon). However, selenium often prefers different coordination geometries, making it less compatible with the tetrahedral building blocks of zeolites.

#### **Possible Alternatives**

- **Gallosilicates:** These are zeolite-like materials where some silicon atoms are replaced by gallium. However, they typically still contain aluminum in their framework.
- **Chalcogenide Zeolite Analogs:** These materials replace the oxygen in traditional zeolites with heavier chalcogen elements like sulfur (S) or selenium (Se). They can incorporate a wider range of metals into their framework.
- **Metal Selenides:** While not true zeolites, certain metal selenides might display porous or zeolite-like structural characteristics.

**Where to Search** To find potentially relevant materials, consider searching these resources:

- **Scientific Databases:** Web of Science, SciFinder, Google Scholar, etc.
- **Crystallographic Databases:** The Cambridge Structural Database (CSD), the Inorganic Crystal Structure Database (ICSD).
- **Materials Science Journals:** Focus on publications within materials chemistry, catalysis, and inorganic chemistry.

Let me know if you'd like me to help you refine your search or explore any of these alternative directions!

### Lack of information response

The response is of low relevance. The chatbot even argues that structures based on Ga and Se atoms do not exist. The response can be accessed via the link: <https://g.co/gemini/share/9681daaf22f2>

## List of Competency Questions

1. What are the channel properties of a specific zeolite formulation?
2. Given a chemical species/counterion, find matching zeolitic frameworks and list their channel properties.
3. Given a chemical species/counterion, find zeolitic frameworks and formulations and filter based on element X, then list characteristic X-ray Diffraction (XRD) signals.
4. Find zeolite materials containing elements X and Y (e.g. Al and P) and list their properties.
5. Given a chemical species/counterion, what is the zeolitic frameworks exhibiting the highest accessible surface area?
6. Plot zeolite frameworks as a function of their occupiable volume and framework density.
7. Show me a crystal structure of a framework that can be described based on natural t-kdq tiling.
8. Show a crystal structure that includes “6-2” secondary building units.
9. Identify a crystal structure that features double zigzag chain structuring.
10. Display the crystallographic structure of a framework containing 'vb 6T' secondary building units (SBUs).
11. How many distinct T-atom positions are there in the -LTA structure, and what is their site multiplicity?
12. Which zeolite framework likely shows  $2\theta$  peaks at 18, 27, and 29 degrees?
13. Which frameworks contain 4\*4\*4 loop configurations, and what are the expected X-ray Diffraction (XRD) peaks for these structures?

### SI.3 Structured Queries and Responses

ZeoliteName	GuestIUPACName	GuestFormula	MolecularWeight	BoilingPoint_value	BoilingPoint_unit
FeAPO-LAU	1<I>H</I>-imidazole	C3H4N2	68.08	530.15	K
Laumontite	oxidane	H2O	18.015	373.124	K
Laumontite	deuterated water	H2O	20.028	374.57	K
Laumontite	potassium(1+)	K+	39.0983	1962.15	K
Laumontite	calcium(2+)	Ca+2	40.08		
Laumontite	sodium(1+)	Na+	22.989769		
Leonhardite	oxidane	H2O	18.015	373.124	K
Leonhardite	potassium(1+)	K+	39.0983	1962.15	K
Leonhardite	calcium(2+)	Ca+2	40.08		
Leonhardite	sodium(1+)	Na+	22.989769		
MnAPO-LAU	1<I>H</I>-imidazole	C3H4N2	68.08	530.15	K
Primary leonhardite	oxidane	H2O	18.015	373.124	K
Primary leonhardite	potassium(1+)	K+	39.0983	1962.15	K
Primary leonhardite	iron(3+)	Fe+3	55.84	1535.0	
Primary leonhardite	calcium(2+)	Ca+2	40.08		
Primary leonhardite	sodium(1+)	Na+	22.989769		
Synthetic laumontite	oxidane	H2O	18.015	373.124	K
Synthetic laumontite	calcium(2+)	Ca+2	40.08		
[Co-Ga-P-O]-LAU	oxidane	H2O	18.015	373.124	K
[Co-Ga-P-O]-LAU	1<I>H</I>-imidazole	C3H4N2	68.08	530.15	K
[Co-Ga-P-O]-LAU	ammonium	H4N+	18.039		
[Fe-Ga-P-O]-LAU	1<I>H</I>-imidazole	C3H4N2	68.08	530.15	K
[Mn-Ga-P-O]-LAU	1<I>H</I>-imidazole	C3H4N2	68.08	530.15	K
[Zn-Ga-P-O]-LAU	1,4-diazabicyclo[2.2.2]octane	C6H12N2	112.17	447.15	K

**Figure S6:** Molecular weights and boiling points of guest species for materials with framework LAU obtained as results of the federated query

### Example queries

**QUERY 1:** Extract crystal lattice parameters with uncertainties (when available)

```

PREFIX zeo: <http://www.theworldavatar.com/kg/ontozeolite/>
PREFIX ocr: <http://www.theworldavatar.com/kg/ontocrystal/>
PREFIX om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>

SELECT ?zeoname ?a ?d_a ?b ?d_b ?c ?d_c ?alpha ?d_alpha ?beta ?d_beta ?gamma ?d_gamma ?volume
↪ ?d_vol ?lattice ?symmNum ?unit_length ?unit_angle

WHERE {
  ?zeo zeo:hasFrameworkCode ?zeoname .
  ?zeo ocr:hasCrystalInformation/ocr:hasUnitCell ?unitcell .

  ?unitcell ocr:hasUnitCellLengths/ocr:hasVectorComponent ?abc_a, ?abc_b, ?abc_c .
  ?abc om:hasUnit/rdfs:label ?unit_length .
  ?abc_a ocr:hasComponentLabel "a"; ocr:hasComponentValue ?a .
  ?abc_b ocr:hasComponentLabel "b"; ocr:hasComponentValue ?b .
  ?abc_c ocr:hasComponentLabel "c"; ocr:hasComponentValue ?c .
  ?unitcell ocr:hasUnitCellAngles/ocr:hasVectorComponent ?abg_a, ?abg_b, ?abg_g .
  ?abg om:hasUnit/rdfs:label ?unit_angle .
  ?abg_a ocr:hasComponentLabel "alpha"; ocr:hasComponentValue ?alpha .
  ?abg_b ocr:hasComponentLabel "beta"; ocr:hasComponentValue ?beta .
  ?abg_g ocr:hasComponentLabel "gamma"; ocr:hasComponentValue ?gamma .

  OPTIONAL { ?unitcell ocr:hasUnitCellVolume/om:hasNumericalValue ?volume . }
  OPTIONAL { ?abc_a ocr:hasComponentUncertainty ?d_a . }
  OPTIONAL { ?abc_b ocr:hasComponentUncertainty ?d_b . }
  OPTIONAL { ?abc_c ocr:hasComponentUncertainty ?d_c . }
  OPTIONAL { ?abg_a ocr:hasComponentUncertainty ?d_alpha . }
  OPTIONAL { ?abg_b ocr:hasComponentUncertainty ?d_beta . }
  OPTIONAL { ?abg_g ocr:hasComponentUncertainty ?d_gamma . }
  OPTIONAL { ?Volume ocr:hasUncertaintyValue ?d_vol . }
  OPTIONAL { ?unitcell ocr:hasLatticeSystem ?lattice . }
  OPTIONAL { ?unitcell ocr:hasSymmetryNumber ?symmNum . }
}

```

**QUERY 2: Extract the coordinate transformation matrix and shift**

**QUERY 3: Extract the atomic coordinates (fractional or Cartesian)**

**QUERY 4: Extract tile information**

**QUERY 5: Extract peaks info of an XRD spectrum**

```

# Fractional to cartesian transformation:
PREFIX zeo: <http://www.theworldavatar.com/kg/ontozeolite/>
PREFIX ocr: <http://www.theworldavatar.com/kg/ontocrystal/>
PREFIX om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>

SELECT ?zeoname ?mc_xx ?mc_xy ?mc_xz ?mc_yx ?mc_yy ?mc_yz ?mc_zx ?mc_zy ?mc_zz ?vc_x ?vc_y ?vc_z
↔ ?mc_unit ?vc_unit

WHERE {
  ?zeo      zeo:hasFrameworkCode ?zeoname .
  ?zeo      ocr:hasCrystalInformation/ocr:hasCoordinateTransformation ?trans.
  ?trans    ocr:hasTransformationMatrixToCartesian ?mCart ;
           ocr:hasTransformationVectorToCartesian ?vCart .

  ?mCart    ocr:hasMatrixComponent ?mc_11, ?mc_12, ?mc_13, ?mc_21, ?mc_22, ?mc_23, ?mc_31,
↔ ?mc_32, ?mc_33 ;    om:hasUnit/rdfs:label ?mc_unit .
  ?mc_11    ocr:hasComponentValue ?mc_xx ;    ocr:hasComponentLabel "xx" .
  ?mc_12    ocr:hasComponentValue ?mc_xy ;    ocr:hasComponentLabel "xy" .
  ?mc_13    ocr:hasComponentValue ?mc_xz ;    ocr:hasComponentLabel "xz" .
  ?mc_21    ocr:hasComponentValue ?mc_yx ;    ocr:hasComponentLabel "yx" .
  ?mc_22    ocr:hasComponentValue ?mc_yy ;    ocr:hasComponentLabel "yy" .
  ?mc_23    ocr:hasComponentValue ?mc_yz ;    ocr:hasComponentLabel "yz" .
  ?mc_31    ocr:hasComponentValue ?mc_zx ;    ocr:hasComponentLabel "zx" .
  ?mc_32    ocr:hasComponentValue ?mc_zy ;    ocr:hasComponentLabel "zy" .
  ?mc_33    ocr:hasComponentValue ?mc_zz ;    ocr:hasComponentLabel "zz" .
  ?vCart    ocr:hasVectorComponent ?vc_1, ?vc_2, ?vc_3 ;    om:hasUnit/rdfs:label ?mc_unit .
  ?vc_1    ocr:hasComponentValue ?vc_x ;    ocr:hasComponentLabel "x" .
  ?vc_2    ocr:hasComponentValue ?vc_y ;    ocr:hasComponentLabel "y" .
  ?vc_3    ocr:hasComponentValue ?vc_z ;    ocr:hasComponentLabel "z" .
}

# Cartesian to fractional transformation:
PREFIX zeo: <http://www.theworldavatar.com/kg/ontozeolite/>
PREFIX ocr: <http://www.theworldavatar.com/kg/ontocrystal/>
PREFIX om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>

SELECT ?zeoname ?mf_xx ?mf_xy ?mf_xz ?mf_yx ?mf_yy ?mf_yz ?mf_zx ?mf_zy ?mf_zz ?vf_x ?vf_y ?vf_z
↔ ?mf_unit ?vf_unit

WHERE {
  ?zeo      zeo:hasFrameworkCode ?zeoname .
  ?zeo      ocr:hasCrystalInformation/ocr:hasCoordinateTransformation ?trans.
  ?trans    ocr:hasTransformationMatrixToCartesian ?mFrac ;
           ocr:hasTransformationVectorToCartesian ?vFrac .

  ?mFrac    ocr:hasMatrixComponent ?mf_11, ?mf_12, ?mf_13, ?mf_21, ?mf_22, ?mf_23, ?mf_31,
↔ ?mf_32, ?mf_33 ;
  ?mf_11    ocr:hasComponentValue ?mf_xx ;    ocr:hasComponentLabel "xx" .
  ?mf_12    ocr:hasComponentValue ?mf_xy ;    ocr:hasComponentLabel "xy" .
  ?mf_13    ocr:hasComponentValue ?mf_xz ;    ocr:hasComponentLabel "xz" .
  ?mf_21    ocr:hasComponentValue ?mf_yx ;    ocr:hasComponentLabel "yx" .
  ?mf_22    ocr:hasComponentValue ?mf_yy ;    ocr:hasComponentLabel "yy" .
  ?mf_23    ocr:hasComponentValue ?mf_yz ;    ocr:hasComponentLabel "yz" .
  ?mf_31    ocr:hasComponentValue ?mf_zx ;    ocr:hasComponentLabel "zx" .
  ?mf_32    ocr:hasComponentValue ?mf_zy ;    ocr:hasComponentLabel "zy" .
  ?mf_33    ocr:hasComponentValue ?mf_zz ;    ocr:hasComponentLabel "zz" .
  ?vFrac    ocr:hasVectorComponent ?vf_1, ?vf_2, ?vf_3 ;    om:hasUnit/rdfs:label ?mf_unit .
  ?vf_1    ocr:hasComponentValue ?vf_x ;    ocr:hasComponentLabel "x" .
  ?vf_2    ocr:hasComponentValue ?vf_y ;    ocr:hasComponentLabel "y" .
  ?vf_3    ocr:hasComponentValue ?vf_z ;    ocr:hasComponentLabel "z" .
}

```



```

PREFIX zeo: <http://www.theworldavatar.com/kg/ontozeolite/>
PREFIX ocr: <http://www.theworldavatar.com/kg/ontocrystal/>
PREFIX om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>

SELECT ?zeoname ?afx ?afy ?afz ?acx ?acy ?acz ?af_unit ?ac_unit ?aoccup ?alabel

WHERE {
  ?zeo zeo:hasFrameworkCode ?zeoname .
  ?zeo ocr:hasCrystalInformation/ocr:hasAtomicStructure/ocr:hasAtomSite ?atom .
  ?atom ocr:hasFractionalPosition ?aF_xyz ; ocr:hasCartesianPosition ?aC_xyz .

  ?aF_xyz ocr:hasVectorComponent ?aF_x, ?aF_y, ?aF_z ; om:hasUnit/rdfs:label ?af_unit .
  ?aF_x ocr:hasComponentValue ?afx ; ocr:hasComponentLabel "x" .
  ?aF_y ocr:hasComponentValue ?afy ; ocr:hasComponentLabel "y" .
  ?aF_z ocr:hasComponentValue ?afz ; ocr:hasComponentLabel "z" .

  ?aC_xyz ocr:hasVectorComponent ?aC_x, ?aC_y, ?aC_z ; om:hasUnit/rdfs:label ?ac_unit .
  ?aC_x ocr:hasComponentValue ?acx ; ocr:hasComponentLabel "x" .
  ?aC_y ocr:hasComponentValue ?acy ; ocr:hasComponentLabel "y" .
  ?aC_z ocr:hasComponentValue ?acz ; ocr:hasComponentLabel "z" .

  OPTIONAL { ?atom ocr:hasAtomSiteLabel ?alabel . }
  OPTIONAL { ?atom ocr:hasOccupancy ?aoccup . }
}

```

```

PREFIX zeo: <http://www.theworldavatar.com/kg/ontozeolite/>
PREFIX ocr: <http://www.theworldavatar.com/kg/ontocrystal/>

SELECT ?zeoname ?nOfTiles ?nOfEdges ?faceNumber ?tileCode ?nFace ?nEdge ?nVert ?tP ?tQ ?tR ?tS
↪ ?signature

WHERE {
  ?zeo zeo:hasFrameworkCode ?zeoname .
  ?zeo ocr:hasCrystalInformation/ocr:hasTiledStructure ?structure .

  ?structure ocr:hasTileNumber ?tileNumber ; ocr:hasTileSignature ?signature ;
  ↪ ocr:hasTransitivity ?trans .

  ?tileNumber ocr:hasValue ?nOfTiles .
  ?tileNumber ocr:isNumberOfTiles ?tile .
  ?tile ocr:hasTileFaceNumber ?tfn .
  ?tile ocr:hasTileCode ?tileCode .
  ?tile ocr:hasNumberOfFaces ?nFace .
  ?tile ocr:hasNumberOfEdges ?nEdge .
  ?tile ocr:hasNumberOfVertices ?nVert .

  ?tfn ocr:isNumberOfTileFaces/ocr:hasNumberOfEdges ?nOfEdges .
  ?tfn ocr:hasValue ?faceNumber .

  ?trans ocr:hasVectorComponent ?transP, ?transQ, ?transR, ?transS .
  ?transP ocr:hasComponentLabel "p" ; ocr:hasComponentValue ?tP .
  ?transQ ocr:hasComponentLabel "q" ; ocr:hasComponentValue ?tQ .
  ?transR ocr:hasComponentLabel "r" ; ocr:hasComponentValue ?tR .
  ?transS ocr:hasComponentLabel "s" ; ocr:hasComponentValue ?tS .
}

```

## SI.4 Description Logic

### TBox for OntoZeolite

The OntoZeolite ontology defines 26 classes, 26 Object Properties, and 28 Data Properties:

```

PREFIX zeo: <http://www.theworldavatar.com/kg/ontozeolite/>
PREFIX ocr: <http://www.theworldavatar.com/kg/ontocrystal/>
PREFIX om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>

SELECT ?zeoname ?2theta ?i ?h ?k ?l ?simulated

WHERE {
  ?zeo zeo:hasFrameworkCode ?zeoname .
  ?zeo ocr:hasCrystalInformation/ocr:hasXRDspectrum/ocr:hasCharacteristicPeak ?peak .
  ?peak ocr:hasTwoThetaPosition ?2theta ; ocr:hasRelativeIntensity ?i .

  ?peak ocr:hasMillerIndices/ocr:hasVectorComponent ?hkl_h, ?hkl_k, ?hkl_l ;
    ocr:isSimulated ?simulated .

  ?hkl_h ocr:hasComponentLabel "h" ; ocr:hasComponentValue ?h .
  ?hkl_k ocr:hasComponentLabel "k" ; ocr:hasComponentValue ?k .
  ?hkl_l ocr:hasComponentLabel "l" ; ocr:hasComponentValue ?l .
}

```

AccessibleAreaPerCell  $\sqsubseteq$  Measure  
 AccessibleAreaPerGram  $\sqsubseteq$  Measure  
 AccessibleVolume  $\sqsubseteq$  Measure  
 AccessibleVolumePerCell  $\sqsubseteq$  Measure  
 CompositeBU  $\sqsubseteq$  MeasureVector  
 CoordinationSequence  $\sqsubseteq$  MeasureVector  
 FrameworkDensity  $\sqsubseteq$  Measure  
 OccupiableAreaPerCell  $\sqsubseteq$  Measure  
 OccupiableAreaPerGram  $\sqsubseteq$  Measure  
 OccupiableVolume  $\sqsubseteq$  Measure  
 OccupiableVolumePerCell  $\sqsubseteq$  Measure  
 Precursor  $\sqsubseteq$  Species  
 RingSizes  $\sqsubseteq$  MeasureVector  
 SecondaryBU  $\sqsubseteq$  MeasureVector  
 SpecificAccessibleArea  $\sqsubseteq$  Measure  
 SpecificOccupiableArea  $\sqsubseteq$  Measure  
 SphereDiameter  $\sqsubseteq$  MeasureVector  
 TAtom  $\sqsubseteq$  Measure  
 VertexSymbol  $\sqsubseteq$  MeasureVector  
 $\exists$ hasAccessibleAreaPerCell.T  $\sqsubseteq$  FrameAccessibleArea  
 $\exists$ hasAccessibleAreaPerCell.T  $\sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasAccessibleAreaPerCell.AccessibleAreaPerCell  
 $\exists$ hasAccessibleAreaPerGram.T  $\sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasAccessibleAreaPerGram.AccessibleAreaPerGram  
 $\exists$ hasAccessibleVolume.T  $\sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasAccessibleVolume.AccessibleVolume  
 $\exists$ hasAccessibleVolumePerCell.T  $\sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasAccessibleVolumePerCell.AccessibleVolumePerCell  
 $\exists$ hasChemicalComposition.T  $\sqsubseteq$  ZeoliticMaterial  
 $\top \sqsubseteq \forall$ hasChemicalComposition.ChemicalComponent

$\exists$ hasCitation.  $\top \sqsubseteq$  Recipe  
 $\top \sqsubseteq \forall$ hasCitation.Document  
 $\exists$ hasCompositeBU.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasCompositeBU.CompositeBU  
 $\exists$ hasCoordinationSequence.  $\top \sqsubseteq$  TAtom  
 $\top \sqsubseteq \forall$ hasCoordinationSequence.CoordinationSequence  
 $\exists$ hasDensity.  $\top \sqsubseteq$  ZeoliticMaterial  
 $\top \sqsubseteq \forall$ hasDensity.Density  
 $\exists$ hasFrameworkDensity.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasFrameworkDensity.FrameworkDensity  
 $\exists$ hasFrameworkTopology.  $\top \sqsubseteq$  ZeoliticMaterial  
 $\top \sqsubseteq \forall$ hasFrameworkTopology.FrameworkTopology  
 $\exists$ hasOccupiableAreaPerCell.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasOccupiableAreaPerCell.OccupiableAreaPerCell  
 $\exists$ hasOccupiableAreaPerGram.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasOccupiableAreaPerGram.OccupiableAreaPerGram  
 $\exists$ hasOccupiableVolume.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasOccupiableVolume.OccupiableVolume  
 $\exists$ hasOccupiableVolumePerCell.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasOccupiableVolumePerCell.OccupiableVolumePerCell  
 $\exists$ hasPrecursor.  $\top \sqsubseteq$  Recipe  
 $\top \sqsubseteq \forall$ hasPrecursor.Precursor  
 $\exists$ hasRecipe.  $\top \sqsubseteq$  ZeoliticMaterial  
 $\top \sqsubseteq \forall$ hasRecipe.Recipe  
 $\exists$ hasRingSizes.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasRingSizes.RingSizes  
 $\exists$ hasSecondaryBU.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasSecondaryBU.SecondaryBU  
 $\exists$ hasSpecificAccessibleArea.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasSpecificAccessibleArea.SpecificAccessibleArea  
 $\exists$ hasSpecificOccupiableArea.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasSpecificOccupiableArea.SpecificOccupiableArea  
 $\exists$ hasSphereDiameter.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasSphereDiameter.SphereDiameter  
 $\exists$ hasTAtom.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasTAtom.TAtom  
 $\exists$ hasTopologicalDensity.  $\top \sqsubseteq$  FrameworkTopology  
 $\top \sqsubseteq \forall$ hasTopologicalDensity.TopologicalDensity  
 $\exists$ hasVertexSymbol.  $\top \sqsubseteq$  TAtom  
 $\top \sqsubseteq \forall$ hasVertexSymbol.VertexSymbol  
 $\exists$ hasZeoliticMaterial.  $\top \sqsubseteq$  ZeoliteFramework  
 $\top \sqsubseteq \forall$ hasZeoliticMaterial.ZeoliticMaterial  $\top \sqsubseteq \forall$ hasAmount.Datatype string  
 $\top \sqsubseteq \forall$ hasBatchComposition.Datatype string  
 $\top \sqsubseteq \forall$ hasBatchPreparation.Datatype string  
 $\top \sqsubseteq \forall$ hasCage.Datatype string  
 $\top \sqsubseteq \forall$ hasChannelDimension.Datatype string

T  $\sqsubseteq$   $\forall$ hasCharacterization.Datatype string  
T  $\sqsubseteq$   $\forall$ hasComposition.Datatype string  
T  $\sqsubseteq$   $\forall$ hasCrystallization.Datatype string  
T  $\sqsubseteq$   $\forall$ hasDOI.Datatype string  
T  $\sqsubseteq$   $\forall$ hasFrameworkCode.Datatype string  
T  $\sqsubseteq$   $\forall$ hasNotes.Datatype string  
T  $\sqsubseteq$   $\forall$ hasProductRecovery.Datatype string  
T  $\sqsubseteq$   $\forall$ hasRCSRLink.Datatype string  
T  $\sqsubseteq$   $\forall$ hasRingCount.Datatype integer  
T  $\sqsubseteq$   $\forall$ hasRingSize.Datatype integer

## TBox for OntoCrystal

The OntoCrystal ontology defines 18 classes, 43 Object Properties, and 25 Data properties:

AbscissaData  $\sqsubseteq$  MeasureVector  
AtomCartesianPosition  $\sqsubseteq$  MeasureVector  
AtomFractionalPosition  $\sqsubseteq$  MeasureVector  
AtomSite  $\sqsubseteq \forall$  hasCartesianPosition.PositionVector  
AtomSite  $\sqsubseteq \forall$  hasElement.OntoSpecies:Species  
AtomSite  $\sqsubseteq \forall$  hasFractionalPosition.PositionVector  
AtomicStructure  $\sqsubseteq \forall$  hasAtomSite.AtomSite  
CoordinateTransformation  $\sqsubseteq \forall$  hasTransformationMatrixToCartesian.TransformationMatrix  
CoordinateTransformation  $\sqsubseteq \forall$  hasTransformationMatrixToFractional.TransformationMatrix  
CoordinateTransformation  $\sqsubseteq \forall$  hasTransformationVectorToCartesian.TransformationVector  
CoordinateTransformation  $\sqsubseteq \forall$  hasTransformationVectorToFractional.TransformationVector  
CrystalInformation  $\sqsubseteq \forall$  hasAtomicStructure.AtomicStructure  
CrystalInformation  $\sqsubseteq \forall$  hasCoordinateTransformation.CoordinateTransformation  
CrystalInformation  $\sqsubseteq \forall$  hasUnitCell.UnitCell  
MatrixComponent  $\sqsubseteq \forall$  hasUnit.om:Unit  
MatrixWithUncertainty  $\sqsubseteq$  MeasureMatrix  
MeasureMatrix  $\sqsubseteq \forall$  hasMatrixComponent.MatrixComponent  
MeasureMatrix  $\sqsubseteq \forall$  hasUnit.om:Unit  
MeasureVector  $\sqsubseteq \forall$  hasUnit.om:Unit  
MeasureVector  $\sqsubseteq \forall$  hasVectorComponent.VectorComponent  
MeasureWithUncertainty  $\sqsubseteq$  om:Measure  
MeasureWithUncertainty  $\sqsubseteq \forall$  hasUnit.om:Unit  
MillerIndices  $\sqsubseteq$  MeasureVector  
OrdinateData  $\sqsubseteq$  MeasureVector  
PositionVector  $\sqsubseteq$  MeasureVector  
TransformationMatrix  $\sqsubseteq$  MatrixWithUncertainty  
TransformationVector  $\sqsubseteq$  VectorWithUncertainty  
Transitivity  $\sqsubseteq$  MeasureVector  
UnitCell  $\sqsubseteq \forall$  hasReciprocalUnitCellAngles.UnitCellAngles  
UnitCell  $\sqsubseteq \forall$  hasReciprocalUnitCellLengths.UnitCellLengths  
UnitCell  $\sqsubseteq \forall$  hasReciprocalUnitCellVectorSet.UnitCellVectorSet  
UnitCell  $\sqsubseteq \forall$  hasUnitCellAngles.UnitCellAngles  
UnitCell  $\sqsubseteq \forall$  hasUnitCellLengths.UnitCellLengths  
UnitCell  $\sqsubseteq \forall$  hasUnitCellVectorSet.UnitCellVectorSet  
UnitCell  $\sqsubseteq \forall$  hasUnitCellVolume.UnitCellVolume  
UnitCellAngles  $\sqsubseteq$  MeasureVector  
UnitCellLatticeVector  $\sqsubseteq$  MeasureVector  
UnitCellLengths  $\sqsubseteq$  MeasureVector

UnitCellVectorSet  $\sqsubseteq \forall$  hasUnitCellVector.UnitCellLatticeVector  
 UnitCellVolume  $\sqsubseteq$  MeasureWithUncertainty  
 VectorComponent  $\sqsubseteq \forall$  hasUnit.om:Unit  
 VectorWithUncertainty  $\sqsubseteq$  MeasureVector  
 $\exists$  hasAbscissaData.  $\top \sqsubseteq$  PlotXY  
 $\exists$  hasAtomSite.  $\top \sqsubseteq$  AtomicStructure  
 $\exists$  hasAtomicStructure.  $\top \sqsubseteq$  CrystalInformation  
 $\exists$  hasAuthor.  $\top \sqsubseteq$  Citation  
 $\exists$  hasAuthorIndex.  $\top \sqsubseteq$  Citation  
 $\exists$  hasCartesianPosition.  $\top \sqsubseteq$  AtomSite  
 $\exists$  hasCharacteristicPeak.  $\top \sqsubseteq$  XRDSpectrum  
 $\exists$  hasCoordinateTransformation.  $\top \sqsubseteq$  CrystalInformation  
 $\exists$  hasElement.  $\top \sqsubseteq$  AtomSite  
 $\exists$  hasFractionalPosition.  $\top \sqsubseteq$  AtomSite  
 $\exists$  hasMatrixComponent.  $\top \sqsubseteq$  MeasureMatrix  
 $\exists$  hasMillerIndices.  $\top \sqsubseteq$  CharacteristicPeak  
 $\exists$  hasOrdinateData.  $\top \sqsubseteq$  PlotXY  
 $\exists$  hasPlotData.  $\top \sqsubseteq$  XRDSpectrum  
 $\exists$  hasPosition.  $\top \sqsubseteq$  TileVertex  
 $\exists$  hasReciprocalUnitCellAngles.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasReciprocalUnitCellLengths.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasReciprocalUnitCellVectorSet.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasTile.  $\top \sqsubseteq$  TiledStructure  
 $\exists$  hasTileFace.  $\top \sqsubseteq$  Tile  
 $\exists$  hasTileFaceNumber.  $\top \sqsubseteq$  Tile  
 $\exists$  hasTileNumber.  $\top \sqsubseteq$  TiledStructure  
 $\exists$  hasTileVertex.  $\top \sqsubseteq$  Tile  
 $\exists$  hasTiledStructure.  $\top \sqsubseteq$  CrystalInformation  
 $\exists$  hasTransformationMatrixToCartesian.  $\top \sqsubseteq$  CoordinateTransformation  
 $\exists$  hasTransformationMatrixToFractional.  $\top \sqsubseteq$  CoordinateTransformation  
 $\exists$  hasTransformationVectorToCartesian.  $\top \sqsubseteq$  CoordinateTransformation  
 $\exists$  hasTransformationVectorToFractional.  $\top \sqsubseteq$  CoordinateTransformation  
 $\exists$  hasTransitivity.  $\top \sqsubseteq$  TiledStructure  
 $\exists$  hasUnit.  $\top \sqsubseteq$  (MatrixComponent  $\sqcup$  MeasureMatrix  $\sqcup$  MeasureVector  $\sqcup$  MeasureWith  
 Uncertainty  $\sqcup$  VectorComponent)  
 $\exists$  hasUnitCell.  $\top \sqsubseteq$  CrystalInformation  
 $\exists$  hasUnitCellAngles.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasUnitCellLengths.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasUnitCellVector.  $\top \sqsubseteq$  UnitCellVectorSet  
 $\exists$  hasUnitCellVectorSet.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasUnitCellVolume.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasVectorComponent.  $\top \sqsubseteq$  MeasureVector  
 $\exists$  hasXRDSpectrum.  $\top \sqsubseteq$  CrystalInformation  
 $\exists$  isAbscissaOf.  $\top \sqsubseteq$  OrdinateData  
 $\exists$  isAuthorIndexOf.  $\top \sqsubseteq$  AuthorIndex  
 $\exists$  isNumberOfTileFaces.  $\top \sqsubseteq$  TileFaceNumber

$\exists$  isNumberOfTiles.  $\top \sqsubseteq$  TileNumber  
 $\top \sqsubseteq \forall$  hasAbscissaData. AbscissaData  
 $\top \sqsubseteq \forall$  hasAtomSite. AtomSite  
 $\top \sqsubseteq \forall$  hasAtomicStructure. AtomicStructure  
 $\top \sqsubseteq \forall$  hasAuthor. Author  
 $\top \sqsubseteq \forall$  hasAuthorIndex. AuthorIndex  
 $\top \sqsubseteq \forall$  hasCartesianPosition. PositionVector  
 $\top \sqsubseteq \forall$  hasCharacteristicPeak. CharacteristicPeak  
 $\top \sqsubseteq \forall$  hasCoordinateTransformation. CoordinateTransformation  
 $\top \sqsubseteq \forall$  hasCrystalInformation. CrystalInformation  
 $\top \sqsubseteq \forall$  hasElement. OntoSpecies:Species  
 $\top \sqsubseteq \forall$  hasFractionalPosition. PositionVector  
 $\top \sqsubseteq \forall$  hasMatrixComponent. MatrixComponent  
 $\top \sqsubseteq \forall$  hasMillerIndices. MillerIndices  
 $\top \sqsubseteq \forall$  hasOrdinateData. OrdinateData  
 $\top \sqsubseteq \forall$  hasPlotData. PlotXY  
 $\top \sqsubseteq \forall$  hasPosition. MeasureVector  
 $\top \sqsubseteq \forall$  hasReciprocalUnitCellAngles. UnitCellAngles  
 $\top \sqsubseteq \forall$  hasReciprocalUnitCellLengths. UnitCellLengths  
 $\top \sqsubseteq \forall$  hasReciprocalUnitCellVectorSet. UnitCellVectorSet  
 $\top \sqsubseteq \forall$  hasTile. Tile  
 $\top \sqsubseteq \forall$  hasTileFace. TileFace  
 $\top \sqsubseteq \forall$  hasTileFaceNumber. TileFaceNumber  
 $\top \sqsubseteq \forall$  hasTileNumber. TileNumber  
 $\top \sqsubseteq \forall$  hasTileVertex. TileVertex  
 $\top \sqsubseteq \forall$  hasTiledStructure. TiledStructure  
 $\top \sqsubseteq \forall$  hasTransformationMatrixToCartesian. TransformationMatrix  
 $\top \sqsubseteq \forall$  hasTransformationMatrixToFractional. TransformationMatrix  
 $\top \sqsubseteq \forall$  hasTransformationVectorToCartesian. TransformationVector  
 $\top \sqsubseteq \forall$  hasTransformationVectorToFractional. TransformationVector  
 $\top \sqsubseteq \forall$  hasTransitivity. Transitivity  
 $\top \sqsubseteq \forall$  hasUnit.om:Unit  
 $\top \sqsubseteq \forall$  hasUnitCell. UnitCell  
 $\top \sqsubseteq \forall$  hasUnitCellAngles. UnitCellAngles  
 $\top \sqsubseteq \forall$  hasUnitCellLengths. UnitCellLengths  
 $\top \sqsubseteq \forall$  hasUnitCellVector. UnitCellLatticeVector  
 $\top \sqsubseteq \forall$  hasUnitCellVectorSet. UnitCellVectorSet  
 $\top \sqsubseteq \forall$  hasUnitCellVolume. UnitCellVolume  
 $\top \sqsubseteq \forall$  hasVectorComponent. VectorComponent  
 $\top \sqsubseteq \forall$  hasXRDSpectrum. XRDSpectrum  
 $\top \sqsubseteq \forall$  isAbscissaOf. AbscissaData  
 $\top \sqsubseteq \forall$  isAuthorIndexOf. Author  
 $\top \sqsubseteq \forall$  isNumberOfTileFaces. TileFace  
 $\top \sqsubseteq \forall$  isNumberOfTiles. Tile  
 $\exists$  <http://purl.org/ontology/bibo/doi>.  $\top \sqsubseteq$  Citation  
 $\exists$  hasAtomSiteLabel.  $\top \sqsubseteq$  AtomSite

$\exists$  hasColumnIndex.  $\top \sqsubseteq$  MatrixComponent  
 $\exists$  hasComponentIndex.  $\top \sqsubseteq$  VectorComponent  
 $\exists$  hasComponentLabel.  $\top \sqsubseteq$  (MatrixComponent  $\sqcup$  VectorComponent)  
 $\exists$  hasComponentUncertainty.  $\top \sqsubseteq$  (MatrixComponent  $\sqcup$  VectorComponent)  
 $\exists$  hasComponentValue.  $\top \sqsubseteq$  (MatrixComponent  $\sqcup$  VectorComponent)  
 $\exists$  hasFaceCode.  $\top \sqsubseteq$  TileFace  
 $\exists$  hasLatticeSystem.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasNumberOfEdges.  $\top \sqsubseteq$  (Tile  $\sqcup$  TileFace)  
 $\exists$  hasNumberOfFaces.  $\top \sqsubseteq$  Tile  
 $\exists$  hasNumberOfVertices.  $\top \sqsubseteq$  Tile  
 $\exists$  hasOccupancy.  $\top \sqsubseteq$  AtomSite  
 $\exists$  hasPeakWidth.  $\top \sqsubseteq$  CharacteristicPeak  
 $\exists$  hasRelativeIntensity.  $\top \sqsubseteq$  CharacteristicPeak  
 $\exists$  hasRowIndex.  $\top \sqsubseteq$  MatrixComponent  
 $\exists$  hasSpaceGroupHM.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasSpaceGroupSymbol.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasSymmetryNumber.  $\top \sqsubseteq$  UnitCell  
 $\exists$  hasTileCode.  $\top \sqsubseteq$  Tile  
 $\exists$  hasTileSignature.  $\top \sqsubseteq$  (Tile  $\sqcup$  TiledStructure)  
 $\exists$  hasTwoThetaPosition.  $\top \sqsubseteq$  CharacteristicPeak  
 $\exists$  hasUncertaintyValue.  $\top \sqsubseteq$  MeasureWithUncertainty  
 $\exists$  hasValue.  $\top \sqsubseteq$  (TileFaceNumber  $\sqcup$  TileNumber)  
 $\exists$  hasVectorLabel.  $\top \sqsubseteq$  UnitCellLatticeVector  
 $\exists$  isSimulated.  $\top \sqsubseteq$  CharacteristicPeak  
 $\exists$  rdfs:label.  $\top \sqsubseteq$  om:Unit  
 $\exists$  http://xmlns.com/foaf/0.1/family\_name.  $\top \sqsubseteq$  Author  
 $\exists$  http://xmlns.com/foaf/0.1/firstName.  $\top \sqsubseteq$  Author  
 $\top \sqsubseteq \forall$  http://purl.org/ontology/bibo/doi.xsd:string  
 $\top \sqsubseteq \forall$  hasAtomSiteLabel.xsd:string  
 $\top \sqsubseteq \forall$  hasColumnIndex.xsd:integer  
 $\top \sqsubseteq \forall$  hasComponentIndex.xsd:integer  
 $\top \sqsubseteq \forall$  hasComponentLabel.xsd:string  
 $\top \sqsubseteq \forall$  hasComponentUncertainty.rdfs:Literal  
 $\top \sqsubseteq \forall$  hasComponentValue.rdfs:Literal  
 $\top \sqsubseteq \forall$  hasFaceCode.xsd:string  
 $\top \sqsubseteq \forall$  hasLatticeSystem.xsd:string  
 $\top \sqsubseteq \forall$  hasNumberOfEdges.xsd:integer  
 $\top \sqsubseteq \forall$  hasNumberOfFaces.xsd:integer  
 $\top \sqsubseteq \forall$  hasNumberOfVertices.xsd:integer  
 $\top \sqsubseteq \forall$  hasOccupancy.xsd:decimal  
 $\top \sqsubseteq \forall$  hasPeakWidth.xsd:decimal  
 $\top \sqsubseteq \forall$  hasRelativeIntensity.xsd:decimal  
 $\top \sqsubseteq \forall$  hasRowIndex.xsd:integer  
 $\top \sqsubseteq \forall$  hasSpaceGroupHM.xsd:string  
 $\top \sqsubseteq \forall$  hasSpaceGroupSymbol.xsd:string  
 $\top \sqsubseteq \forall$  hasSymmetryNumber.xsd:integer



T  $\sqsubseteq$   $\forall$  hasTileCode.xsd:string  
T  $\sqsubseteq$   $\forall$  hasTileSignature.xsd:string  
T  $\sqsubseteq$   $\forall$  hasTwoThetaPosition.xsd:decimal  
T  $\sqsubseteq$   $\forall$  hasUncertaintyValue.xsd:string  
T  $\sqsubseteq$   $\forall$  hasValue.xsd:integer  
T  $\sqsubseteq$   $\forall$  hasVectorLabel.xsd:string  
T  $\sqsubseteq$   $\forall$  isSimulated.xsd:boolean  
T  $\sqsubseteq$   $\forall$  rdfs:label.xsd:string  
T  $\sqsubseteq$   $\forall$  http://xmlns.com/foaf/0.1/family\_name.xsd:string  
T  $\sqsubseteq$   $\forall$  http://xmlns.com/foaf/0.1/firstName.xsd:string