

# Fine-tuned GPT 3.5 for Boosting the Performance of Large Language Models for Question Answering with Knowledge Graph Integration

by Mingze Li

In this file, we need to first process the fine-tuning dataset, then fine-tune the model and finally process the QA tasks with a given dataset contains questions and contexts.

```
In [1]: import tiktoken # for token counting
import numpy as np
import json
from collections import defaultdict
```

## Data analysis for chat model fine-tuning

We have generated the jsonl file for the fine-tuning and we must make sure the dataset can fit the model in the correct format. Those codes for Data Analysis are supported by Data preparation and analysis for chat model fine-tuning:

[https://cookbook.openai.com/examples/chat\\_finetuning\\_data\\_prep](https://cookbook.openai.com/examples/chat_finetuning_data_prep)

```
In [2]: data_path = r"C:\Users\Li\Desktop\fine_tuning_database.jsonl"

# Load the dataset
with open(data_path, 'r', encoding='utf-8') as f:
    dataset = [json.loads(line) for line in f]

# Initial dataset stats
print("Num examples:", len(dataset))
print("First example:")
for message in dataset[0]["messages"]:
    print(message)
```

Num examples: 22

First example:

```
{'role': 'system', 'content': "You are a helpful assistant. Extract and answer using key information from context. Ensure the response is concise, without duplicates, focusing solely on crucial details. There are two examples: Example 1: (Context: The sun is a star in the center of our solar system. Question: What is the sun? Answer: A star at the center of the solar system.) and Example 2: (Context: Neil Armstrong was the first person to walk on the moon. Question: Who was the first person to walk on the moon? Answer: Neil Armstrong.). Your answer must be provided in a direct and concise format, without using any lead-in phrases such as 'Answer:' or similar. Only the answer itself should be included in the response."}
{'role': 'user', 'content': 'Context: Prof. Stefan Diebels has expertise in Computational materials science; PD Dr. Franz Roters has expertise in Computational materials science; EQ2PC has discipline Computational materials science; Prof. Dr. Karsten Durst has expertise in Computational Materials Science; Question: Who is working in the Computational Materials Science field?'}
{'role': 'assistant', 'content': 'Prof. Stefan Diebels; PD Dr. Franz Roters; Prof. Dr. Karsten Durst'}
```

```
In [3]: # Format error checks
format_errors = defaultdict(int)

# Add a list to record the index of the wrong example
missing_assistant_examples = []

for i, ex in enumerate(dataset):
    if not isinstance(ex, dict):
        format_errors["data_type"] += 1
        continue

    messages = ex.get("messages", None)
    if not messages:
        format_errors["missing_messages_list"] += 1
        continue

    for message in messages:
        if "role" not in message or "content" not in message:
            format_errors["message_missing_key"] += 1
            print("errin index:", i)
            print("message: ", message)

        if any(k not in ("role", "content", "name", "function_call") for k in message):
            format_errors["message_unrecognized_key"] += 1
            print("errin index:", i)
            print("message: ", message)

        if message.get("role", None) not in ("system", "user", "assistant", "function_call"):
            format_errors["unrecognized_role"] += 1
            print("errin index:", i)
            print("message: ", message)

        content = message.get("content", None)
        function_call = message.get("function_call", None)

        if (not content and not function_call) or not isinstance(content, str):
            format_errors["missing_content"] += 1
            print("errin index:", i)
            print("message: ", message)

    if not any(message.get("role", None) == "assistant" for message in messages)
```

```

        format_errors["example_missing_assistant_message"] += 1
        missing_assistant_examples.append(i) # 记录发生错误的例子的索引
        print("errin index:", i)

if format_errors:
    print("Found errors:")
    for k, v in format_errors.items():
        print(f"{k}: {v}")
    if missing_assistant_examples:
        print("Missing assistant messages in examples:", missing_assistant_exam
else:
    print("No errors found")

```

No errors found

```

In [4]: encoding = tiktoken.get_encoding("cl100k_base")

# not exact!
# simplified from https://github.com/openai/openai-cookbook/blob/main/examples/H
def num_tokens_from_messages(messages, tokens_per_message=3, tokens_per_name=1):
    num_tokens = 0
    for message in messages:
        num_tokens += tokens_per_message
        for key, value in message.items():
            num_tokens += len(encoding.encode(value))
            if key == "name":
                num_tokens += tokens_per_name
    num_tokens += 3
    return num_tokens

def num_assistant_tokens_from_messages(messages):
    num_tokens = 0
    for message in messages:
        if message["role"] == "assistant":
            num_tokens += len(encoding.encode(message["content"]))
    return num_tokens

def print_distribution(values, name):
    print(f"\n#### Distribution of {name}:")
    print(f"min / max: {min(values)}, {max(values)}")
    print(f"mean / median: {np.mean(values)}, {np.median(values)}")
    print(f"p5 / p95: {np.quantile(values, 0.1)}, {np.quantile(values, 0.9)}")

```

```

In [5]: # Warnings and tokens counts
n_missing_system = 0
n_missing_user = 0
n_messages = []
convo_lens = []
assistant_message_lens = []

for ex in dataset:
    messages = ex["messages"]
    if not any(message["role"] == "system" for message in messages):
        n_missing_system += 1
    if not any(message["role"] == "user" for message in messages):
        n_missing_user += 1
    n_messages.append(len(messages))
    convo_lens.append(num_tokens_from_messages(messages))
    assistant_message_lens.append(num_assistant_tokens_from_messages(messages))

```

```

print("Num examples missing system message:", n_missing_system)
print("Num examples missing user message:", n_missing_user)
print_distribution(n_messages, "num_messages_per_example")
print_distribution(convo_lens, "num_total_tokens_per_example")
print_distribution(assistant_message_lens, "num_assistant_tokens_per_example")
n_too_long = sum(l > 4096 for l in convo_lens)
print(f"\n{n_too_long} examples may be over the 4096 token limit, they will be t

```

Num examples missing system message: 0

Num examples missing user message: 0

#### Distribution of num\_messages\_per\_example:

min / max: 3, 3

mean / median: 3.0, 3.0

p5 / p95: 3.0, 3.0

#### Distribution of num\_total\_tokens\_per\_example:

min / max: 254, 3463

mean / median: 570.5454545454545, 390.5

p5 / p95: 280.6, 690.0000000000001

#### Distribution of num\_assistant\_tokens\_per\_example:

min / max: 2, 97

mean / median: 18.227272727272727, 8.5

p5 / p95: 3.0, 32.7

0 examples may be over the 4096 token limit, they will be truncated during fine-tuning

```

In [6]: # Pricing and default n_epochs estimate
MAX_TOKENS_PER_EXAMPLE = 4096

TARGET_EPOCHS = 3
MIN_TARGET_EXAMPLES = 100
MAX_TARGET_EXAMPLES = 25000
MIN_DEFAULT_EPOCHS = 1
MAX_DEFAULT_EPOCHS = 25

n_epochs = TARGET_EPOCHS
n_train_examples = len(dataset)
if n_train_examples * TARGET_EPOCHS < MIN_TARGET_EXAMPLES:
    n_epochs = min(MAX_DEFAULT_EPOCHS, MIN_TARGET_EXAMPLES // n_train_examples)
elif n_train_examples * TARGET_EPOCHS > MAX_TARGET_EXAMPLES:
    n_epochs = max(MIN_DEFAULT_EPOCHS, MAX_TARGET_EXAMPLES // n_train_examples)

n_billing_tokens_in_dataset = sum(min(MAX_TOKENS_PER_EXAMPLE, length) for length in dataset)
print(f"Dataset has ~{n_billing_tokens_in_dataset} tokens that will be charged for during training")
print(f"By default, you'll train for {n_epochs} epochs on this dataset")
print(f"By default, you'll be charged for ~{n_epochs * n_billing_tokens_in_dataset} tokens")

```

Dataset has ~12552 tokens that will be charged for during training

By default, you'll train for 4 epochs on this dataset

By default, you'll be charged for ~50208 tokens

## Fine tuning

```

In [7]: file_path = data_path

```

## upload the fine tuning dataset

In this step we need to collect key variables for the fine-tuned model:

- 1.file\_object\_id
- 2.fine\_tuning\_job\_id
- 3.fine\_tuned\_model\_name

Note: If you want to fit the model with new datasets please open those blocks  
blew and reset key variables

```
In [8]: """# open if you want to upload a new dataset

from openai import OpenAI
client = OpenAI()
# use English version: converted_messages_en.jsonl
file_object = client.files.create(
    file=open(file_path, "rb"),# could be: #messages.jsonl,#converted_messages.jsonl
    purpose="fine-tune"
)
print("file_object.id:",file_object.id)
file_object_id = file_object.id
file_object

# open when you want to upload new data for fine-tuning"""
```

file\_object.id: file-HadPRSrLyP6QNu8PD5ES0xkj

```
Out[8]: FileObject(id='file-HadPRSrLyP6QNu8PD5ES0xkj', bytes=57902, created_at=17094865
99, filename='fine_tuning_database.jsonl', object='file', purpose='fine-tune',
status='processed', status_details=None)
```

```
In [24]: file_object_id = 'file-RFWxvsMTKaJuqT9JP5wxusip' #'file-MCDZL5LcwJ8fVLS5VnK9x5oz'
```

## Fine-tuning model

fine\_tuning\_job fine\_tuning\_job.id="ftjob-YtoxfhaeCv5EDCjMt1PGbvHZ"

fine\_tuned\_model\_name = "gpt-3.5-turbo-0613"

```
In [10]: """# open if you want to upload a new dataset
# open when you want to upload new data for fine-tuning
from openai import OpenAI
client = OpenAI()

fine_tuning_job = client.fine_tuning.jobs.create(
    training_file = file_object_id,
    model="gpt-3.5-turbo"
)
print(fine_tuning_job.id)
fine_tuning_job

#FineTuningJob(id='ftjob-2Lpyr3aaKh1qmUa2Piq0r9ma', created_at=1704830407, error
""")
```

ftjob-WMVtrj0tCvTHA6F2H2M7Ugkj

```
Out[10]: FineTuningJob(id='ftjob-WMVtrj0tCvTHA6F2H2M7Ugkj', created_at=1709486605, error=Error(code=None, message=None, param=None, error=None), fine_tuned_model=None, finished_at=None, hyperparameters=Hyperparameters(n_epochs='auto', batch_size='auto', learning_rate_multiplier='auto'), model='gpt-3.5-turbo-0125', object='fine_tuning.job', organization_id='org-1RBrqOHK4MGbSBFmx0Tqvb1b', result_files=[], status='validating_files', trained_tokens=None, training_file='file-MCDZL5lcwJ8fVLS5VnK9x5oz', validation_file=None, user_provided_suffix=None)
```

```
In [11]: fine_tuning_job_id = 'ftjob-0hMtkRxKZX1W1K0vNPaVsAKA'#"ftjob-WMVtrj0tCvTHA6F2H2M7Ugkj"
```

```
In [12]: """# open if you want to upload a new dataset
from openai import OpenAI
import time

client = OpenAI()

# Loop to check the status of the fine-tuning job
while True:
    fine_tuning_job = client.fine_tuning.jobs.retrieve(fine_tuning_job_id)
    if fine_tuning_job.status == 'succeeded':
        # The fine-tuning job is completed and the name of the fine-tuned model
        fine_tuned_model_name = fine_tuning_job.fine_tuned_model
        print("finetuned model name:", fine_tuned_model_name)
        break
    elif fine_tuning_job.status == 'failed':
        print("Fine-tuning job failed.")
        break
    print("Wait for the fine-tuning job to complete...")
    time.sleep(60)
"""
```

```
Wait for the fine-tuning job to complete...
Wait for the fine-tuning job to complete...
Wait for the fine-tuning job to complete...
Wait for the fine-tuning job to complete...
Wait for the fine-tuning job to complete...
finetuned model name: ft:gpt-3.5-turbo-0125:personal::8yjjvAGTt
```

```
In [25]: fine_tuned_model_name = "ft:gpt-3.5-turbo-0125:personal::8yR07PVm" #"ft:gpt-3.5-turbo-0125:personal::8yjjvAGTt"
```

## Load the dataset contains questions and contexts

If you have some new datasets please reload them here. Note: the number of questions must not less than the number of context

```
In [26]: import pandas as pd
path = r"C:\Users\Li\Desktop\train_8_without score.xlsx" #r"C:\Users\Li\Desktop\train_8_without score.xlsx"
df = pd.read_excel(path)
df
```

Out[26]:

	Competency Question	Ground Truth	
0	Who is working in the Computational Materials ...	PD Dr. habil. Thomas Hammerschmidt; Prof. Dr. ...	kar
1	What are the research projects associated to E...	VIMMP (2018-2021); OYSTER (2017-2021); SimDOME...	kar
2	Who are the contributors of the data "datasets"?	Prof. Felix Fritzen < <a href="http://demo.fiz-karlsruhe...">http://demo.fiz-karlsruhe...</a>	kar
3	Who is working with Researcher "Ebrahim Norouz..."	Prof. Dr. Harald Sack; Mirza Mohtashim Alam; D...	kar
4	who is the email address of "ParaView"?	support@kitware.com	[('https://schema.org
5	What are the affiliations of Volker Hofmann?	Forschungszentrum Jülich < <a href="http://demo.fiz-karl...">http://demo.fiz-karl...</a>	kar
6	What is "Molecular Dynamics" Software? List th...	1. Resource: <a href="http://demo.fiz-karlsruhe.de/matw...">http://demo.fiz-karlsruhe.de/matw...</a>	kar
7	What are pre- and post-processing tools for MD...	Pizza.py Toolkit; pycsal; ASE; MDTraj; freud	kar
8	What are some workflow environments for comput...	Pyiron; AiiDA; SimStack	kar
9	How should I cite pyiron?	"title = {pyiron: An integrated development en...	kar
10	Where can I find a list of interatomic potenti...	Interatomic potentials repository (NIST); Open...	kar
11	What are python libraries used	Pyscal	kar

	Competency Question	Ground Truth	
	for calculating...		
12	What are the electronic lab notebooks available?	Pasta ELN; eLabFTW; Karlsruhe Data Infrastruct...	[('https://w3id.org/sk
13	What are the software for Molecular Dynamics (...)	Calphy; Melting temperature computational work...	kar
14	What are the ontologies in nanomaterials domain?	MaterialsMine (MM); NanoParticle Ontology for ...	kar
15	What is DAMASK?	DAMASK is a unified multi-physics crystal plas...	kar
16	What are the data portals for materials scienc...	MatPortal	kar
17	What are the instruments for APT?	No Answer	[('n80e2298a9637
18	In which institution can I find tomography equ...	1. Procon CT-alpha"\n "X-ray micro computed to...	
19	What are the educational resources for Ontology?	Knowledge Graphs (openHPI Massive Open Online ...	kar
20	What is the API of Materials Project?	https://api.materialsproject.org/docs	kar
21	Which simulation software have a python API?	EMMOntoPy	kar
22	What is the documentation of the "MatDB Online"?	https://publications.jrc.ec.europa.eu/reposito...	kar



	Competency Question	Ground Truth	
23	What are the types of software licenses?	BSD-3-Clause license; GPL 3 License; ASL/ BSD-...	
24	What are the software used to produce the data...	pymatgen; AiiDA; Quantum Espresso	kar
25	What are datasets produced by the BAM organiza...	BAM reference data: results of ASTM E139 -11 c...	kar
26	What are some available datasets of mechanical...	Full dataset of several mechanical tests on an...	kar
27	What are datasets related to "Transmission ele...	Radii of S-phase Al <sub>2</sub> CuMg in Al-alloy EN AW-261...	kar
28	What is the license of the dataset "Elastic Co...	BSD 3-Clause License	kar
29	What is the repository for "BAM reference data"?	<a href="http://demo.fiz-karlsruhe.de/matwerk/E291704/">http://demo.fiz-karlsruhe.de/matwerk/E291704/</a> ; ...	kar
30	What are the different data formats in the "BA...	csv; pdf	kar
31	What is the software version of "pacemaker"?	0.2.7	kar
32	What is the field of research "BAM reference d...	high temperature mechanical behavior	kar
33	What is the description of the "BAM reference ...	The creep behavior of a certified reference ma...	kar
34	What are the datasets	Brinell-Hardness (HBW 2.5/62.5) of Al-alloy EN...	

	Competency Question	Ground Truth
	produced in 2022?	
35	Who is the creators of the "BAM reference data"?	Sina Schriever ( <a href="http://demo.fiz-karlsruhe.de/m...">http://demo.fiz-karlsruhe.de/m...</a> ) kar
36	What are the datasets published by "BAM"?	Datasets for the analysis of dislocations at g... kar
37	Average	NaN

```
In [27]: questions = df['Competency Question']
contexts = df['Context']
length = questions.count()
print(questions.count())
print(contexts.count())
```

38

35

## QA tasks based on the fine-tuned model

```
In [30]: import os
from openai import OpenAI
import openai

def get_answer_with_single_question(question, context):

    # Set OpenAI API key
    api_key = os.environ.get('OPENAI_API_KEY')
    openai.api_key = api_key

    # Initialize OpenAI client
    client = OpenAI(api_key=api_key)

    # Set up the model
    model= fine_tuned_model_name #'gpt-3.5-turbo-1106'

    try:
        messages=[
            {"role": "system", "content": "You are a helpful assistant."},
            # instruction
            {"role": "user", "content": "Context: " + context + " Question: " + question}
        ]
        response = client.chat.completions.create(
            model=model,
            messages=messages
        )
```

```

        # Extract and return the answer
        answer = response.choices[0].message.content
        return answer

    except Exception as e:
        print(f"An error occurred while processing the problem: {e}")
        return "Unable to get answer"

```

```

In [31]: from openai import OpenAI
import openai
import os

# Set OpenAI API key
api_key = os.environ.get('OPENAI_API_KEY')
openai.api_key = api_key

# Initialize OpenAI client
client = OpenAI(api_key=api_key)

# Set up the model
model = fine_tuned_model_name #'gpt-3.5-turbo-1106'

user_input_count = 0
arr_answers = []

for i in range(length):
    if contexts[i] == "":
        user_input_count += 1
        continue
    print(f"No. {user_input_count} question: {questions[i]}")
    print(f"No. {user_input_count} context: {contexts[i]}")
    answer = get_answer_with_single_question(questions[i], contexts[i])
    arr_answers.append(answer)
    print(f"No. {user_input_count} answer: {answer}")
    user_input_count += 1

```

No. 0 question: Who is working in the Computational Materials Science field?

No. 0 context: Pyiron YouTube channel has description "Welcome to our youtube channel where we present pyiron - an integrated development environment for computational materials science. "; MaterialsMine (MM) has description "A materials ontology to support data publication involving nanomaterials and metamaterials."; Materials Design Ontology (MDO) has related Project Open Databases Integration for Materials Design (OPTIMADE); Materials Design Ontology (MDO) has documentation [w3id.org/mdo/full/1.0/](https://www.mdo.org/mdo/full/1.0/); Materials Design Ontology (MDO) has description "MDO is an ontology for materials design field, representing the domain knowledge specifically related to solid-state physics and computational materials science."; Open Materials Database has description "The Open Materials Database is a publicly available, searchable database for information on the properties of materials created by people who work in computational materials design research. "; Prof. Dr. Karsten Durst has expertise in Computational Materials Science; Prof. Dr.-Ing. Stephan Wulfinghoff has expertise in Computational Materials Science; Christian Dorn has expertise in Computational Materials Science; Dr.-Ing Abril Azocar Guzman has expertise in Computational Materials Science; Prof. Dr. Jörg Neugebauer has expertise in Computational Materials Science; pyscal\_rdf has discipline Computational Materials Science; Elastic Constant Demo has discipline Computational Materials Science; Computational Material Sample Ontology has discipline Computational Materials Science; Elastic Constant Demo Data has discipline Computational Materials Science; MaterialsProject has description "By computing properties of all known materials, the Materials Project aims to remove guesswork from materials design in a variety of applications. Experimental research can be targeted to the most promising compounds from computational data sets. Researchers will be able to data-mine scientific trends in materials properties. By providing materials researchers with the information they need to design better, the Materials Project aims to accelerate innovation in materials research."; Avogadro has documentation <https://avogadro.cc/docs/>; Avogadro has description "molecule editor and visualizer designed for cross-platform use in computational chemistry, molecular modeling, bioinformatics, materials science, and related areas"; Polymer Genome has description "An informatics platform for polymer property prediction and design using machine learning"; Prof. Stefan Diebels has expertise in Computational materials science; PD Dr. Franz Roters has expertise in Computational materials science; EQ2PC has discipline Computational materials science; Thermo-Calc has discipline computational materials science; Thermo-Calc has discipline thermodynamics; Thermo-Calc has description "Thermo-Calc, is used by materials scientists and engineers to generate material properties data, gain insights about materials, understand a specific observation, and answer direct questions related to a specific material and/or its processing. Used in conjunction with suitable databases, Thermo-Calc can be used for a wide variety of applications."; PD Dr. habil. Thomas Hammerschmidt has expertise in computational materials science; Dr Sarath Menon has expertise in computational materials science; Calphy has discipline computational materials science; Melting temperature computational workflow has discipline computational materials science; Pyscal has discipline computational materials science; MinimumEnergyPoints has discipline computational materials science; Image based prediction of the heat conduction tensor has discipline computational materials science; Finite Element Analysis Program has discipline computational materials science; Vienna Ab initio Simulation Package has discipline computational materials science; Cambridge Serial Total Energy Package has discipline computational materials science; Carr Parrinello Molecular Dynamics has discipline computational materials science; ABINIT has discipline computational materials science; BigDFT has discipline computational materials science; Parallel total energy has discipline computational materials science; JDFTx has discipline computational materials science; PARSEC has discipline computational materials science; CP2K has discipline computational materials science; GPAW has discipline computational materials science; S/PHI/nX has discipline computational materials science; Qbox First-Principles Molecular Dynamics has discipline computational materials science; DFTK.jl has discipline computational materials science; density of Montréal has discipline computational materials science; SIESTA has discipline computational materials science; CRYSTAL

has discipline computational materials science; FHI-AIMS has discipline computational materials science; FPLO has discipline computational materials science; Open source package for Material eXplorer has discipline computational materials science; Elk has discipline computational materials science; exciting has discipline computational materials science; FLEUR has discipline computational materials science; WIEN2k has discipline computational materials science; Large-scale Atomic/Molecular Massively Parallel Simulator has discipline computational materials science; The ITAP Molecular Dynamics Program has discipline computational materials science; GROMACS has discipline computational materials science; MD++ has discipline computational materials science; Thermo-Calc has discipline computational materials science; OpenPhase has discipline computational materials science; Multiphysics Object Oriented Simulation Environment has discipline computational materials science; Abaqus has discipline computational materials science; FeniCS Project has discipline computational materials science; Dusseldorf Advanced Material Simulation Kit has discipline computational materials science; Automated interactive infrastructure and database for computational science has discipline computational materials science; SimStack has discipline computational materials science; Atomic Simulation Environment has discipline computational materials science; Quantum Espresso has discipline computational materials science; Pyiron has discipline computational materials science; TURBOMOLE has discipline Computational Material Science

No. 0 answer: Prof. Dr. Karsten Durst; Prof. Dr.-Ing. Stephan Wulfinghoff; Christian Dorn; Dr.-Ing. Abril Azocar Guzman; Prof. Dr. Jörg Neugebauer; Prof. Stefan Diebels; PD Dr. Franz Roters; PD Dr. habil. Thomas Hammerschmidt; Dr. Sarath Menon

No. 1 question: What are the research projects associated to EMMO?

No. 1 context: Elemental Multiperspective Material Ontology (EMMO) has related Project EMMC-CSA (2016-2019); Elemental Multiperspective Material Ontology (EMMO) has related Project SimDOME (2019-2023); Elemental Multiperspective Material Ontology (EMMO) has related Project MarketPlace (2018-2022); Elemental Multiperspective Material Ontology (EMMO) has related Project VIMMP (2018-2021); Elemental Multiperspective Material Ontology (EMMO) has related Project OntoTrans (2020-2024); Elemental Multiperspective Material Ontology (EMMO) has related Project ReaxPro (2019-2023); Elemental Multiperspective Material Ontology (EMMO) has related Project OntoCommons (2020-2023); Elemental Multiperspective Material Ontology (EMMO) has related Project OYSTER (2017-2021); Elemental Multiperspective Material Ontology (EMMO) has related Project NanoMECommons (2021-2025); Elemental Multiperspective Material Ontology (EMMO) has related Project OpenModel (2021-2025); Elemental Multiperspective Material Ontology (EMMO) has description "The Elemental Multiperspective Material Ontology (EMMO) is the result of a multidisciplinary effort within the EMMC, aimed at the development of a standard representational ontology framework based on current materials modelling and characterization knowledge."; Essential Source of Schemas and Examples (ESSE) has description "JSON schemas and examples representing structural data, characteristic properties, modeling workflows and related data about materials standardizing the diverse landscape of information"

No. 1 answer: Project EMMC-CSA (2016-2019); SimDOME (2019-2023); MarketPlace (2018-2022); VIMMP (2018-2021); OntoTrans (2020-2024); ReaxPro (2019-2023); OntoCommons (2020-2023); OYSTER (2017-2021); NanoMECommons (2021-2025); OpenModel (2021-2025)

No. 2 question: Who are the contributors of the data "datasets"?

No. 2 context: Materials Data Repository has contributor National Institute for Materials Science (NIMS), Japan; Materials Data Repository has description "MDR is a data repository to collect and store papers, presentation materials, and related materials data to accumulate and release them in a form suitable for the promotion of materials research and materials informatics. Users can search the documents and the data from information (metadata) such as sample, instrument, method, and from the full text of the deposited data, to browse and download them freely."; Phonon database at Kyoto university has description "It's getting to be a mainstream idea collecting a big set of ab-initio calculations and analyzing them by informatics approach. For scientific purpose, we plan to do it for phonon property"

ies. As an attempt, we have run a set of ab-initio phonon calculations and left the raw data on this static web site. The aim of this web site is just leaving the data to distribute."; datasets has contributor Fernández; datasets has contributor Prof. Felix Fritzen; datasets has contributor Oliver Weeger; datasets has contributor Frederik Larsson; datasets has description "This dataset was used in the work "Material modeling for parametric, anisotropic finite hyperelasticity based on machine learning with application in optimization of metamaterials" by M. Fernández, F. Fritzen and O. Weeger"; datasets has description "This dataset was used in the work "On-the-Fly Adaptivity for Nonlinear Twoscale Simulations Using Artificial Neural Networks and Reduced Order Modeling" by F. Fritzen, M. Fernández, and F. Larsson"

No. 2 answer: Fernández; Prof. Felix Fritzen; Oliver Weeger; Frederik Larsson

No. 3 question: Who is working with Researcher "Ebrahim Norouzi" in the same group?

No. 3 context: Ebrahim Norouzi has work package FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis has work package FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has work package FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has work package FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam has expertise in Knowledge Graph Embedding; Mirza Mohtashim Alam has expertise in Machine Learning; Mirza Mohtashim Alam has expertise in NLP; Mirza Mohtashim Alam has expertise in Computer Science; Ebrahim Norouzi has expertise in FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis has expertise in FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has expertise in FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has expertise in FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi has funding project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis has funding project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has funding project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has funding project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi related participant project is FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis related participant project is FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl related participant project is FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack related participant project is FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi has related Project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis has related Project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has related Project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has related Project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi has discipline FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis has discipline FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has discipline FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has discipline FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi has parent organisation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis has parent organisation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has parent organisation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has parent organisation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi has related task area FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis has related task area FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has related task area FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has related task area FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi has documentation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. J

örg Waitelonis has documentation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has documentation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has documentation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Jörg Waitelonis has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Dr. Heike Fliegl has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Prof. Dr. Harald Sack has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; PD Dr. Steffen Brinckmann has work package Forschungszentrum Jülich; Dr. Volker Hofmann has work package Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has work package Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has work package Forschungszentrum Jülich; Said Fathalla has work package Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has work package Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has work package Forschungszentrum Jülich; Ahmad Zainul Ihsan has expertise in Materials Data Science and Informatics; PD Dr. Steffen Brinckmann has expertise in Forschungszentrum Jülich; Dr. Volker Hofmann has expertise in Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has expertise in Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has expertise in Forschungszentrum Jülich; Said Fathalla has expertise in Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has expertise in Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has expertise in Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has funding project Forschungszentrum Jülich; Dr. Volker Hofmann has funding project Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has funding project Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has funding project Forschungszentrum Jülich; Said Fathalla has funding project Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has funding project Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has funding project Forschungszentrum Jülich; PD Dr. Steffen Brinckmann related participant project is Forschungszentrum Jülich; Dr. Volker Hofmann related participant project is Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman related participant project is Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger related participant project is Forschungszentrum Jülich; Said Fathalla related participant project is Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko related participant project is Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld related participant project is Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has related Project Forschungszentrum Jülich; Dr. Volker Hofmann has related Project Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has related Project Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has related Project Forschungszentrum Jülich; Said Fathalla has related Project Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has related Project Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has related Project Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has discipline Forschungszentrum Jülich; Dr. Volker Hofmann has discipline Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has discipline Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has discipline Forschungszentrum Jülich; Said Fathalla has discipline Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has discipline Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has discipline Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has parent organisation Forschungszentrum Jülich; Dr. Volker Hofmann has parent organisation Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has parent organisation Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has parent organisation Forschungszentrum Jülich; Said Fathalla has parent organisation Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has parent organisation Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has parent organisation Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has related task area Forschungszentrum Jülich; Dr. Volker Hofmann has related task area Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has related task area Forschungszentrum Jülich; Prof. Dr. Ruth Schw

aiger has related task area Forschungszentrum Jülich; Said Fathalla has related task area Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has related task area Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has related task area Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has documentation Forschungszentrum Jülich; Dr. Volker Hofmann has documentation Forschungszentrum Jülich; Dr.-Ing. Abril Azocar Guzman has documentation Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has documentation Forschungszentrum Jülich; Said Fathalla has documentation Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has documentation Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has documentation Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has description Forschungszentrum Jülich; Dr. Volker Hofmann has description Forschungszentrum Jülich; Dr.-Ing. Abril Azocar Guzman has description Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has description Forschungszentrum Jülich; Said Fathalla has description Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has description Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has description Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has description Forschungszentrum Jülich; Dr. Volker Hofmann has description Forschungszentrum Jülich; Dr.-Ing. Abril Azocar Guzman has description Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has description Forschungszentrum Jülich; Said Fathalla has description Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has description Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has description Forschungszentrum Jülich; Phonon database at Kyoto university has description "It's getting to be a mainstream idea collecting a big set of ab-initio calculations and analyzing them by informatics approach. For scientific purpose, we plan to do it for phonon properties. As an attempt, we have run a set of ab-initio phonon calculations and left the raw data on this static web site. The aim of this web site is just leaving the data to distribute."; Mirza Mohtashim Alam has work package FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Ebrahim Norouzi has expertise in Information Service Engineering; Ebrahim Norouzi has expertise in Materials Informatic; Mirza Mohtashim Alam has expertise in FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam has funding project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam related participant project is FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam has related Project FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam has discipline FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam has parent organisation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam has related task area FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam has documentation FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Mirza Mohtashim Alam has description FIZ Karlsruhe - Leibniz Institute for Information Infrastructure; Thomas Pardoen has work package Institute of Mechanics; Thomas Pardoen has expertise in Institute of Mechanics; Thomas Pardoen has funding project Institute of Mechanics; Thomas Pardoen related participant project is Institute of Mechanics; Thomas Pardoen has related Project Institute of Mechanics; Thomas Pardoen has discipline Institute of Mechanics; Thomas Pardoen has parent organisation Institute of Mechanics; Thomas Pardoen has related task area Institute of Mechanics; Thomas Pardoen has documentation Institute of Mechanics; Thomas Pardoen has description Institute of Mechanics; Thomas Pardoen has description Institute of Mechanics

No. 3 answer: Mirza Mohtashim Alam

No. 4 question: who is the email address of "ParaView"?

No. 4 context: dateCreated has comment "The date on which the CreativeWork was created or the item was added to a DataFeed."; dateCreated has domain resource; dataset has comment "A dataset is associated with a structured information about a resource."; about material has domain dataset; has creator has domain dataset; ParaView has website <https://www.paraview.org/>; ParaView has contact point support@kitware.com; ParaView has repository <https://gitlab.kitware.com/paraview/paraview>; ParaView has description "ParaView is an open-source, multi-platform data analysis and visualization application based on Visualization Toolkit (VTK)."; data por



tal has comment "An online platform that provides access to data collections and datasets."

No. 4 answer: support@kitware.com

No. 5 question: What are the affiliations of Volker Hofmann?

No. 5 context: Dr. Tilmann Hickel has affiliation with Max-Planck-Institut für Eisenforschung; Dr. Tilmann Hickel has affiliation with Bundesanstalt für Materialforschung und -prüfung; PD Dr. Franz Roters has affiliation with Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has affiliation with Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has affiliation with Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has affiliation with Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters has curation status Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has curation status Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has curation status Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has curation status Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters has parent organisation Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has parent organisation Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has parent organisation Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has parent organisation Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters has citation Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has citation Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has citation Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has citation Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters has documentation Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has documentation Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has documentation Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has documentation Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters related participant project is Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek related participant project is Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer related participant project is Max-Planck-Institut für Eisenforschung; Dr Sarath Menon related participant project is Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters has elucidation Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has elucidation Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has elucidation Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has elucidation Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters has definition Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has definition Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has definition Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has definition Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters has definition Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has definition Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has definition Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has definition Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters has description Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has description Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has description Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has description Max-Planck-Institut für Eisenforschung; PD Dr. Franz Roters has description Max-Planck-Institut für Eisenforschung; Prof. Dr. Erik Bitzek has description Max-Planck-Institut für Eisenforschung; Prof. Dr. Jörg Neugebauer has description Max-Planck-Institut für Eisenforschung; Dr Sarath Menon has description Max-Planck-Institut für Eisenforschung; Niklas Siemer has affiliation with Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer has affiliation with Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer has curation status Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer has parent organisation Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer has citation Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer has documentation Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer related participant project is Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer has elucidation Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer has definition Max-Planck-Institut für Eisenforschung GmbH

H; Prof. Dr. Jörg Neugebauer has definition Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer has description Max-Planck-Institut für Eisenforschung GmbH; Prof. Dr. Jörg Neugebauer has description Max-Planck-Institut für Eisenforschung GmbH; Dr. Volker Hofmann has affiliation with Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has affiliation with Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has affiliation with Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has affiliation with Forschungszentrum Jülich; Said Fathalla has affiliation with Forschungszentrum Jülich; Ahmad Zainul Ihsan has affiliation with Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has affiliation with Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has affiliation with Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has curation status Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has curation status Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has curation status Forschungszentrum Jülich; Said Fathalla has curation status Forschungszentrum Jülich; Ahmad Zainul Ihsan has curation status Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has curation status Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has curation status Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has parent organisation Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has parent organisation Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has parent organisation Forschungszentrum Jülich; Said Fathalla has parent organisation Forschungszentrum Jülich; Ahmad Zainul Ihsan has parent organisation Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has parent organisation Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has parent organisation Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has citation Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has citation Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has citation Forschungszentrum Jülich; Said Fathalla has citation Forschungszentrum Jülich; Ahmad Zainul Ihsan has citation Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has citation Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has citation Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has documentation Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has documentation Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has documentation Forschungszentrum Jülich; Said Fathalla has documentation Forschungszentrum Jülich; Ahmad Zainul Ihsan has documentation Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has documentation Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has documentation Forschungszentrum Jülich; PD Dr. Steffen Brinckmann related participant project is Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman related participant project is Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger related participant project is Forschungszentrum Jülich; Said Fathalla related participant project is Forschungszentrum Jülich; Ahmad Zainul Ihsan related participant project is Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko related participant project is Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld related participant project is Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has elucidation Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has elucidation Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has elucidation Forschungszentrum Jülich; Said Fathalla has elucidation Forschungszentrum Jülich; Ahmad Zainul Ihsan has elucidation Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has elucidation Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has elucidation Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has definition Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has definition Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has definition Forschungszentrum Jülich; Said Fathalla has definition Forschungszentrum Jülich; Ahmad Zainul Ihsan has definition Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has definition Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has definition Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has definition Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has definition Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has definition Forschungszentrum Jülich; Said Fathalla has definition Forschungszentrum Jülich; Ahmad Zainul Ihsan has definition Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has definition Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has definition Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has description Forschungszentrum Jülich; Dr.-Ing Abril Azocar Guzman has description Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has description Forschungszentrum Jülich; Said Fathalla

la has description Forschungszentrum Jülich; Ahmad Zainul Ihsan has description Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has description Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has description Forschungszentrum Jülich; PD Dr. Steffen Brinckmann has description Forschungszentrum Jülich; Dr.-Ing. Abril Azocar Guzman has description Forschungszentrum Jülich; Prof. Dr. Ruth Schwaiger has description Forschungszentrum Jülich; Said Fathalla has description Forschungszentrum Jülich; Ahmad Zainul Ihsan has description Forschungszentrum Jülich; Dr.-Ing. Hanna Tsybenko has description Forschungszentrum Jülich; Prof. Dr. Stefan Sandfeld has description Forschungszentrum Jülich

No. 5 answer: Forschungszentrum Jülich

No. 6 question: What is "Molecular Dynamics" Software? List the programming language, documentation page, repository, and license information.

No. 6 context: OpenBIS has description "The openBIS platform has three primary functionalities: 1. Inventory management of laboratory samples, materials, protocols, equipment. 2. Laboratory notebook, to document lab experiments. 3. Data management, to store all data related to lab experiments (raw, processed, analysed data, scripts, Jupyter notebooks, etc.)."; PubChem has citation Kim S. Exploring Chemical Information in PubChem. Curr. Protoc.; 2021 Aug 9; 1(8):e217. doi: <https://doi.org/10.1002/cpz1.217>; PubChem has description "PubChem is the world's largest collection of freely accessible chemical information. Search chemicals by name, molecular formula, structure, and other identifiers. Find chemical and physical properties, biological activities, safety and toxicity information, patents, literature citations and more."; Graphical Interface for Materials Simulations has description "browser-based toolbox for electronic structure codes and supports the generation of input files for first-principles electronic structure calculations and workflows, as well as the analysis and visualization of the resulting data extracted from the output files."; Visual Molecular Dynamics has documentation <http://www.ks.uiuc.edu/Research/vmd/current/ug/>; Visual Molecular Dynamics has description "molecular modelling and visualization computer program"; Physical Information File (PIF) has documentation <https://citrineinformatics.github.io/pif-documentation/index.html>; Physical Information File (PIF) has description "The Physical Information File (PIF) is a schema that is designed to impose structure on information about physical systems. Schema design, as well as tools for working with PIF-formatted files using the python (pypif) and java (jpif) programming languages are maintained by Citrine Informatics."; Computational Materials Repository (CMR) has related resource ASE; Computational Materials Repository (CMR) has description "Each CMR project consists of an ASE-database and a project page describing the data and showing examples of how to work with the data using Python and ASE."; GROMACS has citation \* GROMACS: A message-passing parallel molecular dynamics implementation\_x000D\_

H.J.C. Berendsen, D. van der Spoel and R. van Drunen\_x000D\_

Comp. Phys. Comm. 91, 43-56 (1995)\_x000D\_

DOI: [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)\_x000D\_

\_x000D\_

\* GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable\_x000D\_ molecular simulation\_x000D\_

B. Hess and C. Kutzner and D. van der Spoel and E. Lindahl\_x000D\_

J. Chem. Theory Comput. 4 (2008) pp. 435-447\_x000D\_

DOI: <https://doi.org/10.1021/ct700301q>\_x000D\_

\_x000D\_

\* GROMACS 4.5: a high-throughput and highly parallel open source\_x000D\_ molecular simulation toolkit\_x000D\_

Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar\_x000D\_

Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson\_x000D\_

David van der Spoel, Berk Hess, Erik Lindahl\_x000D\_

Bioinformatics 29 (2013) pp. 845-54\_x000D\_

DOI: <https://doi.org/10.1093/bioinformatics/btt055>\_x000D\_

\_x000D\_

\* Tackling Exascale Software Challenges in Molecular Dynamics Simulations\_x000D\_ with GROMACS\_x000D\_

Szilárd Páll, Mark J. Abraham, Carsten Kutzner, Berk Hess, Erik Lindahl  
In S. Markidis & E. Laure (Eds.), Solving Software Challenges for Exascale,

Lecture Notes for Computer Science, 8759 (2015) pp. 3-27

DOI: [https://doi.org/10.1007/978-3-319-15976-8\\_1](https://doi.org/10.1007/978-3-319-15976-8_1)

\* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers

M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl,

SoftwareX, 1, (2015), 19-25

DOI: <https://doi.org/10.1016/j.softx.2015.06.001>; Cluster Expansion in Atomic Simulation Environment has documentation [https://computationalmaterials.gitlab.io/cluster/api\\_doc.html](https://computationalmaterials.gitlab.io/cluster/api_doc.html); Cluster Expansion in Atomic Simulation Environment has citation J. Chang, D. Kleiven, M. Melander, J. Akola, J. M. Garcia-Lastra and T. Vegge: A versatile and user-friendly implementation of Cluster Expansion method Journal of Physics: Condensed Matter; Cluster Expansion in Atomic Simulation Environment has description "package that automates the cumbersome setup and construction procedure of cluster expansion (CE). It provides a comprehensive list of tools for specifying parameters for CE, generating training structures, fitting effective cluster interaction (ECI) values and running Monte Carlo simulations."

No. 6 answer: GROMACS; Programming Language: C, C++; Documentation: DOI links provided; Repository: maintained by developers; License: GNU Lesser General Public License

No. 7 question: What are pre- and post-processing tools for MD simulations?

No. 7 context: Graphical Interface for Materials Simulations has operating system operating system; Graphical Interface for Materials Simulations has description "browser-based toolbox for electronic structure codes and supports the generation of input files for first-principles electronic structure calculations and workflows, as well as the analysis and visualization of the resulting data extracted from the output files."; Extensible Self-optimizing Phase Equilibria Infrastructure has description "tool for creating CALPHAD databases and evaluating the uncertainty of CALPHAD models."; AML has description "Python package to automatically build the reference set for the training of Neural Network Potentials (NNPs), and eventually other machine-learned potentials, in an automated, data-driven fashion."; freud has documentation <https://freud.readthedocs.io/en/latest/>; freud has description "The freud Python library provides a simple, flexible, powerful set of tools for analyzing trajectories obtained from molecular dynamics or Monte Carlo simulations."; MDTraj has description "Read, write and analyze MD trajectories with only a few lines of Python code."; Crystallography Open Database (COD) has description "Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers."; OVITO basic has documentation <https://www.ovito.org/docs/current/>; OVITO basic has operating system Linux; OVITO basic has operating system Mac OS; OVITO basic has operating system Microsoft Windows; OVITO basic has description "visualization and analysis for data generated in MD, atomistic Monte-Carlo and other particle-based simulations"; Pizza.py Toolkit has documentation <https://lammps.github.io/pizza/doc/Manual.html>; Pizza.py Toolkit has description "Pizza.py is a loosely integrated collection of tools, many of which provide pre- and post-processing capability for the LAMMPS molecular dynamics, ChemCell cell modeling, and SPPARKS kinetic Monte Carlo packages. There are tools to create input files, convert between file formats, process log and dump files, create plots, and visualize and animate simulation snapshots."; Yet another force field has documentation <http://molmod.github.io/yaff/>; Yet another force field has description "pythonic force-field (FF) code used at the Center for Molecular Modeling (CMM) to test-drive new FF models"

No. 7 answer: Pizza.py Toolkit

No. 8 question: What are some workflow environments for computational materials science?

No. 8 context: matminer has description "Python library for data mining the properties of materials."; Simmate has description "The Simulated Materials Ecosystem"

m (Simmate) is a toolbox and framework for computational materials research. It lets you explore various crystal databases, predict new materials, and quickly calculate properties."; freud has description "The freud Python library provides a simple, flexible, powerful set of tools for analyzing trajectories obtained from molecular dynamics or Monte Carlo simulations. "; The Materials Data Facility (MDF) has description "A simple way to publish, discover, and access materials datasets"; MGI JSON Schema has description "The purpose of this repository is to enable community development of JSON schema within materials science and engineering."; Polymer Genome has description "An informatics platform for polymer property prediction and design using machine learning"; AtomisticTools has related resource Pyiron; Pyiron has description "pyiron is an integrated development environment for implementing, testing, and running simulations in computational materials science."; pycalphad has description "Python library for computational thermodynamics using the CALPHAD method."

No. 8 answer: matminer, Simmate, Pyiron

No. 9 question: How should I cite pyiron?

No. 9 context: APT analysis using CompositionSpace and pyiron has description "Workflow to analyse APT data using CompositionSpace jobs implemented in pyiron"; Pyrho has description "Pyrho is a real-space DFT code written in Python. It is not built to be super-fast or scalable- instead, it is built to be super-readable. There's a tutorial notebook (tutorial.ipynb) to guide you through the process of building Pyrho from scratch (starting with the 1D Schrodinger equation). This notebook is based on interactive lectures given by Prof. Joerg Neugebauer."; Pyiron has description "pyiron is an integrated development environment for implementing, testing, and running simulations in computational materials science."; Pyiron has citation title = {pyiron: An integrated development environment for computational materials science},\_x000D\_

journal = {Computational Materials Science},\_x000D\_

volume = {163},\_x000D\_

pages = {24 - 36},\_x000D\_

year = {2019},\_x000D\_

issn = {0927-0256},\_x000D\_

doi = {https://doi.org/10.1016/j.commatsci.2018.07.043},\_x000D\_

url = {http://www.sciencedirect.com/science/article/pii/S0927025618304786},\_x000D\_

author = {Jan Janssen and Sudarsan Surendralal and Yury Lysogorskiy and Mira Todorova and Tilmann Hickel and Ralf Drautz and Jörg Neugebauer},\_x000D\_

keywords = {Modelling workflow, Integrated development environment, Complex simulation protocols},

No. 9 answer: Jan Janssen and colleagues.

No. 10 question: Where can I find a list of interatomic potentials?

No. 10 context: mendeleev has description "A python package for accessing various properties of elements, ions and isotopes in the periodic table of elements."; mendeleev has documentation <https://mendeleev.readthedocs.io/en/stable/quick.html>; Interatomic potentials repository (NIST) has citation C.A. Becker, F. Tavazza, Z.T. Trautt, and R.A. Buarque de Macedoc (2013), "Considerations for choosing and using force fields and interatomic potentials in materials science and engineering," Current Opinion in Solid State and Materials Science, 17, 277-283. DOI: 10.1016/j.cossms.2013.10.001 L.M. Hale, Z.T. Trautt, and C.A. Becker (2018), "Evaluating variability with atomistic simulations: the effect of potential and calculation methodology on the modeling of lattice and elastic constants," Modelling and Simulation in Materials Science and Engineering, 26, 055003. DOI: 10.1088/1361-651X/aabc05; pacemaker has description "a tool for fitting of interatomic potentials in a general nonlinear Atomic Cluster Expansion (ACE) form. "; pacemaker has documentation <https://pacemaker.readthedocs.io/en/latest/>; pacemaker has citation Bochkarev, A., Lysogorskiy, Y., Menon, S., Qamar, M., Mrovec, M., & Drautz, R. (2022). Efficient parametrization of the atomic cluster expansion. Physical Review Materials, 6(1), 013804\_x000D\_

Lysogorskiy, Y., Oord, C. V. D., Bochkarev, A., Menon, S., Rinaldi, M., Hammerschmidt, T., ... & Drautz, R. (2021). Performant implementation of the atomic cluster

r expansion (PACE) and application to copper and silicon. npj Computational Materials, 7(1), 1-12; NGLview has description "Jupyter widget to interactively view molecular structures and trajectories"; NGLview has citation Hai Nguyen, David A Case, Alexander S Rose; NGLview - Interactive molecular graphics for Jupyter notebooks, Bioinformatics, , btx789, <https://doi.org/10.1093/bioinformatics/btx789>; The ITAP Molecular Dynamics Program has description "IMD is a software package for classical molecular dynamics simulations. Several types of interactions are supported, such as central pair potentials, EAM potentials for metals, Stillinger-Weber and Tersoff potentials for covalent systems, and Gay-Berne potentials for liquid crystals."; The ITAP Molecular Dynamics Program has documentation <http://imd.itap.physik.uni-stuttgart.de/userguide/imd.html>; The ITAP Molecular Dynamics Program has citation J. Roth, F. Gähler, and H.-R. Trebin, [x000D\\_](#)  
A molecular dynamics run with 5.180.116.000 particles, [x000D\\_](#)  
Int. J. Mod. Phys. C 11, 317-22 (2000). [x000D\\_](#)  
PDF-File (184k) [x000D\\_](#)  
[x000D\\_](#)  
J. Stadler, R. Mikulla, and H.-R. Trebin, [x000D\\_](#)  
IMD: A Software Package for Molecular Dynamics Studies on Parallel Computers, [x000D\\_](#)  
[0D\\_](#)  
Int. J. Mod. Phys. C 8, 1131-1140 (1997). [x000D\\_](#)  
gzipped PostScript (690k) PDF-File (325k) [x000D\\_](#)  
[x000D\\_](#)  
J. Roth [x000D\\_](#)  
IMD: A Typical Massively Parallel Molecular Dynamics Code for Classical Simulations - Structure, Applications, Latest Developments [x000D\\_](#)  
Sustained Simulation Performance 2013, [x000D\\_](#)  
eds. M. M. Resch, W. Bez, E. Forcht, H. Kobayashi, Y. Kovalenko, Springer, Heidelberg 2013, pp. 63-76.; Image based prediction of the heat conduction tensor has description "An open-source code providing a graphical user interface to predict the effective heat conductivity of microstructures and requires only the image data as input. The deployed methods are proposed and validated in the related paper, doi: 10.3390/mca24020057."; Image based prediction of the heat conduction tensor has citation Data-Driven Microstructure Property Relations; The Science Hub for Atomic-scale Research at Chalmers has description "Dielectric functions of metallic alloys calculated with time-dependent density functional theory. Atoms under pressure - Electron configuration and electronegativity of the atoms under compression. Wulff construction - The classic method for calculating equilibrium nanoparticle shapes in a modern setting. SQS - Generate special quasirandom atomic structures with a few clicks in your browser."; Workshop: Workflows for atomistic simulations has description "Workshop lectures and videos from fitting of interatomic potentials to validating them."; Interatomic potentials repository (NIST) has description "This repository provides a source for interatomic potentials (force fields), related files, and evaluation tools to help researchers obtain interatomic models and judge their quality and applicability. "; Interatomic potentials repository (NIST) has citation C.A. Becker, F. Tavazza, Z.T. Trautt, and R.A. Buehler (2013), "Considerations for choosing and using force fields and interatomic potentials in materials science and engineering," Current Opinion in Solid State and Materials Science, 17, 277-283. DOI: 10.1016/j.cossms.2013.10.001  
L.M. Hale, Z.T. Trautt, and C.A. Becker (2018), "Evaluating variability with atomistic simulations: the effect of potential and calculation methodology on the modeling of lattice and elastic constants," Modelling and Simulation in Materials Science and Engineering, 26, 055003. DOI: 10.1088/1361-651X/aabc05  
No. 10 answer: Interatomic potentials repository (NIST)  
No. 11 question: What are python libraries used for calculating local atomic structural environment?  
No. 11 context: Materials Commons has repository <https://github.com/materials-commons/materialscommons.org>; Materials Commons has description "A site for Materials Scientists to collaborate, store and publish research. The Materials Commons is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering."; matminer has repository

hub.com/hackingmaterials/matminer; matminer has description "Python library for data mining the properties of materials."; Pyscal has repository <https://github.com/pyscal/pyscal>; Pyscal uses method molecular dynamics; Pyscal uses method atomistic simulations; Pyscal has description "Python library written in C++ for calculation of local atomic structural environment"; Materials Cloud has repository <https://github.com/materialscloud-org>; Materials Cloud Youtube channel required to ol Materials Cloud; Materials Cloud has description "Materials Cloud is built to enable the seamless sharing and dissemination of resources in computational materials science, offering educational, research, and archiving tools; simulation software and services; and curated and raw data. These underpin published results and empower data-based discovery, compliant with data management plans and the FAIR principles."; Calphy has repository <https://github.com/ICAMS/calphy>; Calphy uses method molecular dynamics; Calphy uses method atomistic simulations; Calphy has description "Python library and command line tool for calculation of free energies"; Data Processing for Engineers and Scientists required tool python; Selected Chapters in Data Processing: Microstructure Analysis and Synthesis required tool python; Information Service Engineering required tool python; EQ2PC has repository <https://github.com/DataAnalyticsEngineering/EQ2PC>; EQ2PC has description "A python script to implement the generation of discrete periodic structures with identical 2-point correlation."; OpenPathSampling has repository <https://github.com/openpathsampling/openpathsampling>; OpenPathSampling has description "Python library to facilitate path sampling algorithms."; py4vasp has repository <https://github.com/orest-d/p4vasp>; py4vasp has description "Python interface to extract data from VASP calculations"; icet has repository <https://gitlab.com/materials-modeling/icet>; icet has description "icet is a python tool for the construction and sampling of alloy cluster expansions."; Cluster Expansion in Atomic Simulation Environment has repository <https://gitlab.com/computationalmaterials/cluster-expansion>; Cluster Expansion in Atomic Simulation Environment has description "package that automates the cumbersome setup and construction procedure of cluster expansion (CE). It provides a comprehensive list of tools for specifying parameters for CE, generating training structures, fitting effective cluster interaction (ECI) values and running Monte Carlo simulations."

No. 11 answer: Pyscal

No. 12 question: What are the electronic lab notebooks available?

No. 12 context: Laboratory Instrument has type nan; electronic lab notebook has type nan; Atomic Simulation Environment has type software; Atomic Simulation Environment has type processing software; Atomic Simulation Environment has type research software; Atomic Simulation Environment has type visualization software; Atomic Simulation Environment uses method atomistic simulations; Atomic Simulation Environment has description "set of tools and Python modules for setting up, manipulating, running, visualizing and analyzing atomistic simulations"; Jupyter notebooks has type object type; Datasets for the analysis of dislocations at grain boundaries and during vein formation in cyclically deformed Ni micropillars has object type Jupyter notebooks; Experimental Workflow has type nan

No. 12 answer: nan

No. 13 question: What are the software for Molecular Dynamics (MD)?

No. 13 context: matminer has website <https://hackingmaterials.lbl.gov/matminer/#>; matminer has programming Language python; matminer has description "Python library for data mining the properties of materials."; FitSNAP has website <https://hackingmaterials.lbl.gov/matminer/>; FitSNAP has programming Language python; FitSNAP has programming Language cython; FitSNAP has description "A Python Package For Training SNAP Interatomic Potentials for use in the LAMMPS molecular dynamics package"; dynasor has website <https://dynasor.materialsmodeling.org/>; dynasor has description "dynasor is a simple tool for calculating total and partial dynamical structure factors as well as current correlation functions from molecular dynamics (MD) simulations."; Simmate has website <https://simmate.org/>; Simmate has description "The Simulated Materials Ecosystem (Simmate) is a toolbox and framework for computational materials research. It lets you explore various crystal databases, predict new materials, and quickly calculate properties."; Visual Molecular Dynamics has operating system Linux; Visual Molecular Dynamics has operating system

m Mac OS; Visual Molecular Dynamics has operating system Microsoft Windows; Visual Molecular Dynamics has website <https://www.ks.uiuc.edu/Research/vmd/>; Visual Molecular Dynamics has description "molecular modelling and visualization computer program"; freud has programming Language C++; freud has programming Language python; freud has description "The freud Python library provides a simple, flexible, powerful set of tools for analyzing trajectories obtained from molecular dynamics or Monte Carlo simulations. "; Vienna Ab initio Simulation Package has website <http://www.vasp.at/>; Vienna Ab initio Simulation Package has programming Language Fortran; Vienna Ab initio Simulation Package has description "The Vienna Ab initio Simulation Package (VASP) is a computer program for atomic scale materials modelling, e.g. electronic structure calculations and quantum-mechanical molecular dynamics, from first principles."; QuickFF has website <https://molmod.github.io/QuickFF/>; QuickFF has programming Language python; QuickFF has description "Python package developed at the Center for Molecular Modeling (CMM) to quickly derive accurate force fields from ab initio calculations. "; Crystallography Open Database (COD) has website <http://www.crystallography.net/cod/>; Crystallography Open Database (COD) has description "Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers."; Yet another force field has website <https://molmod.ugent.be/software/yaff/>; Yet another force field has programming Language python; Yet another force field has programming Language C; Yet another force field has description "pythonic force-field (FF) code used at the Center for Molecular Modeling (CMM) to test-drive new FF models"

No. 13 answer: dynasor; freud

No. 14 question: What are the ontologies in nanomaterials domain?

No. 14 context: Knowledge Graphs (openHPI Massive Open Online Course (MOOC)) has description "In this course you will learn what is necessary to design, implement, and use knowledge graphs. The focus of this course will be on basic semantic technologies including the principles of knowledge representation and symbolic AI. This includes information encoding via RDF triples, knowledge representation via ontologies with OWL, efficiently querying knowledge graphs via SPARQL, latent representation of knowledge in vector space, as well as knowledge graph applications in innovative information systems, as e.g., semantic and exploratory search."; MaterialsMine (MM) has description "A materials ontology to support data publication involving nanomaterials and metamaterials."; MaterialsMine (MM) has semantic expressivity nan; Materials Design Ontology (MDO) has description "MDO is an ontology for materials design field, representing the domain knowledge specifically related to solid-state physics and computational materials science."; Materials Design Ontology (MDO) has semantic expressivity nan; Materials And Molecules Basic Ontology (MAMBO) has description "MAMBO (Materials And Molecules Basic Ontology) is a domain ontology for molecular materials. "; Materials And Molecules Basic Ontology (MAMBO) has semantic expressivity nan; Essential Source of Schemas and Examples (ESSE) has description "JSON schemas and examples representing structural data, characteristic properties, modeling workflows and related data about materials standardizing the diverse landscape of information"; eNanoMapper ontology has description "The eNanoMapper ontologies aim to provide a comprehensive suite of ontologies for the nanomaterial safety assessment domain"; eNanoMapper ontology has semantic expressivity nan; SimPhoNy has description "Simulation framework for multi-scale phenomena in micro- and nanosystems"; Computational Material Sample Ontology has description "CMSO is an ontology that aims to describe computational materials science samples (or structures), including crystalline defects."; Computational Material Sample Ontology has semantic expressivity nan

No. 14 answer: MaterialsMine (MM)

Materials Design Ontology (MDO)

Materials And Molecules Basic Ontology (MAMBO)

eNanoMapper ontology

No. 15 question: What is DAMASK?

No. 15 context: Dusseldorf Advanced Material Simulation Kit has type software; Dusseldorf Advanced Material Simulation Kit has type research software; Dusseldorf Advanced Material Simulation Kit uses method crystal plasticity; Dusseldorf Advanced



ced Material Simulation Kit has description "DAMASK is a unified multi-physics crystal plasticity simulation package. The solution of continuum mechanical boundary value problems requires a constitutive response that connects deformation and stress at each material point. This problem is solved in DAMASK on the basis of crystal plasticity using a variety of constitutive models and homogenization approaches. However, treating mechanics in isolation is no longer sufficient to study emergent advanced high-strength materials. In these materials, deformation happens interrelated with displacive phase transformation, significant heating, and potential damage evolution. Therefore, DAMASK is capable of handling multi-physics problems. Following a modular approach, additional field equations are solved in a fully coupled way using a staggered approach.

"; Elk has type software; Elk has type research software; Elk uses method Density functional theory; Elk has description "An all-electron full-potential linearised augmented-plane wave (LAPW) code with many advanced features. Written originally at Karl-Franzens-Universität Graz as a milestone of the EXCITING EU Research and Training Network, the code is designed to be as simple as possible so that new developments in the field of density functional theory (DFT) can be added quickly and reliably. "; NOMAD has type funding project; Novel Materials Discovery (NOMAD) has type data portal; Novel Materials Discovery (NOMAD) has description "The Novel Materials Discovery (NOMAD) Laboratory maintains the largest worldwide repository for input and output files of all important computational materials science computer programs."

No. 15 answer: A unified multi-physics crystal plasticity simulation package

No. 16 question: What are the data portals for materials science ontologies?

No. 16 context: MaterialsMine (MM) has repository <https://github.com/tetherless-world/materialsmine>; MaterialsMine (MM) has description "A materials ontology to support data publication involving nanomaterials and metamaterials."; Materials Design Ontology (MDO) has repository <https://github.com/LiUSemWeb/Materials-Design-Ontology>; Materials Design Ontology (MDO) has description "MDO is an ontology for materials design field, representing the domain knowledge specifically related to solid-state physics and computational materials science."; Materials And Molecules Basic Ontology (MAMBO) has repository <https://github.com/daimoners/MAMBO>; Materials And Molecules Basic Ontology (MAMBO) has description "MAMBO (Materials And Molecules Basic Ontology) is a domain ontology for molecular materials. "; Simmate has repository <https://github.com/jacksund/simmate>; Simmate has website <https://simmate.org/>; Simmate has description "The Simulated Materials Ecosystem (Simmate) is a toolbox and framework for computational materials research. It lets you explore various crystal databases, predict new materials, and quickly calculate properties."; MatPortal has website <https://matportal.org/>; MatPortal has description "Ontology repository for materials science."; Essential Source of Schemas and Examples (ESSE) has repository <https://github.com/Exabyte-io/esse>; Essential Source of Schemas and Examples (ESSE) has description "JSON schemas and examples representing structural data, characteristic properties, modeling workflows and related data about materials standardizing the diverse landscape of information"; MGI JSON Schema has repository <https://github.com/usnistgov/mgi-json-schema>; MGI JSON Schema has description "The purpose of this repository is to enable community development of JSON schema within materials science and engineering."; Development of coupled ontologies and workflows for thermochemical treatments has main task area Ontologies for Materials Science; Alignment of application- and higher-level ontologies has main task area Ontologies for Materials Science; Co-creation environment for experts has main task area Ontologies for Materials Science; Adaptive automated characterization pipelines and meta data schemas for high throughput experiments has main task area Ontologies for Materials Science; Method- and scale-bridging workflows and data structures for tomography has main task area Ontologies for Materials Science; Unified ontology for matrix-inclusion microstructure and composites has main task area Ontologies for Materials Science; Ontologies for defects in crystals has main task area Ontologies for Materials Science; Ontologies for Materials Science has description "This TA develops Ontologies and metadata schemes in close cooperation with the stakeholders, tailored to their resp

ective needs."; Computational Material Sample Ontology has repository <https://github.com/Materials-Data-Science-and-Informatics/cmso-ontology>; Computational Material Sample Ontology has contact point Dr.-Ing Abril Azocar Guzman; Computational Material Sample Ontology has description "CMSO is an ontology that aims to describe computational materials science samples (or structures), including crystalline defects."

No. 16 answer: MatPortal

No. 17 question: What are the instruments for APT?

No. 17 context: :n80e2298a963740ea9bdf3f1a27d3951b27 has first element Instrument; APT analysis using CompositionSpace and pyiron has operating system Unix; APT analysis using CompositionSpace and pyiron has operating system Linux; APT analysis using CompositionSpace and pyiron has operating system Mac OS; APT analysis using CompositionSpace and pyiron has description "Workflow to analyse APT data using CompositionSpace jobs implemented in pyiron"; Crystallography Open Database (COD) has citation Gražulis, S., Daškevič, A., Merkys, A., Chateigner, D., Lutterotti, L., Quirós, M., Serebryanaya, N. R., Moeck, P., Downs, R. T. & LeBail, A. (2012). Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. Nucleic Acids Research, 40, D420-D427. doi: 10.1093/nar/gkr900 (BibTeX, EndNote/Refer, plain text)

Grazulis, S., Chateigner, D., Downs, R. T., Yokochi, A. T., Quiros, M., Lutterotti, L., Manakova, E., Butkus, J., Moeck, P. & Le Bail, A. (2009). Crystallography Open Database – an open-access collection of crystal structures. Journal of Applied Crystallography, 42, 726-729. doi: 10.1107/S0021889809016690 (BibTeX, EndNote/Refer, plain text); Crystallography Open Database (COD) has description "Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers."

No. 17 answer: Instrument

No. 18 question: In which institution can I find tomography equipment?

No. 18 context: nan

An error occurred while processing the problem: can only concatenate str (not "float") to str

No. 18 answer: Unable to get answer

No. 19 question: What are the educational resources for Ontology?

No. 19 context: MatPortal has description "Ontology repository for materials science."; Dr. Amir Laadhar has expertise in ontology matching; Additive Manufacturing Ontology (AMONTOLOGY) has description "The AM ontology has been developed following two major milestones. The ontology developed within the first milestone includes AMProcessOntology, ModelOntology and AMOntology files. AMProcessOntology contains the set of entities used to capture knowledge about additive manufacturing processes. ModelOntology contains the set of entities used to capture knowledge about modeling concepts that represent (possibly) multi-physics multi-scale processes. AMOntology uses AMProcessOntology and ModelOntology files to describe entities that capture knowledge about characteristics of computational models for AM processes."; Ontologies for Materials Science has description "This TA develops Ontologies and metadata schemes in close cooperation with the stakeholders, tailored to their respective needs."; The MatWerk ontology has description "The MatWerk ontology represents research data and related activities of the MSE community. A first version of the ontology will be simplified, focusing on (i) community structure: researchers, research projects, universities, and institutions; (ii) infrastructure: software, workflows, ontologies, schemas, APIs, instruments, facilities, educational resources; and (iii) data: repositories, databases, scientific publications, published datasets and reference data."

No. 19 answer: Infrastructure: software, workflows, ontologies, schemas, APIs, instruments, facilities, educational resources

No. 20 question: What is the API of Materials Project?

No. 20 context: Python Materials Genomics has repository <https://github.com/materialsproject/pymatgen>; Python Materials Genomics has website <https://pymatgen.org/>; Python Materials Genomics has description "materials analysis code that defines core object representations for structures and molecules with support for man

y electronic structure codes. It is currently the core analysis code powering the Materials Project."; Materials Platform for Data Science has website <https://mpd.s.io/>; Materials Platform for Data Science has description "Online materials data base (known as PAULING FILE project) with nearly 2 million entries: physical properties, crystal structures, phase diagrams, available via API, ready for modern data-intensive applications. The source of these entries are about 300,000 peer-reviewed publications in materials science, processed during the last 16 years by an international team of PhD editors. The results are presented online with a quick search interface. The basic access is provided for free."; Materials Cloud has repository <https://github.com/materialscloud-org>; Materials Cloud has website <https://www.materialscloud.org/discover/mc3d/dashboard/ptable>; Materials Cloud has description "Materials Cloud is built to enable the seamless sharing and dissemination of resources in computational materials science, offering educational, research, and archiving tools; simulation software and services; and curated and raw data. These underpin published results and empower data-based discovery, compliant with data management plans and the FAIR principles."; Hybrid<sup>3</sup> materials database has website <https://materials.hybrid3.duke.edu/>; Hybrid<sup>3</sup> materials database has description "The Hybrid<sup>3</sup> materials database provides a comprehensive collection of experimental and computational materials data for crystalline organic-inorganic compounds, predominantly based on the perovskite paradigm."

No. 20 answer: <https://pymatgen.org/>

No. 21 question: Which simulation software have a python API?

No. 21 context: matminer has description "Python library for data mining the properties of materials."; matminer has citation Ward, L., Dunn, A., Faghaninia, A., Zimmermann, N. E. R., Bajaj, S., Wang, Q., Montoya, J. H., Chen, J., Bystrom, K., Dylla, M., Chard, K., Asta, M., Persson, K., Snyder, G. J., Foster, I., Jain, A., Matminer: An open source toolkit for materials data mining. Comput. Mater. Sci. 152, 60-69 (2018).; OPTIMADE has citation Andersen et al, OPTIMADE, an API for exchanging materials data, Sci. Data 8, 217 (2021) 10.1038/s41597-021-00974-z; Pyscal has description "Python library written in C++ for calculation of local atomic structural environment"; Pyscal has documentation <https://pyscal.org>; Pyscal has citation pyscal: A python module for structural analysis of atomic environments.; CompositionSpace has description "CompositionSpace is a python library for analysis of APT data. "; CompositionSpace has documentation <https://compositionsphere.readthedocs.io/en/latest/>; pycp2k has description "Python Cp2k interface"; OpenPathSampling has description "Python library to facilitate path sampling algorithms."; OpenPathSampling has citation David W.H. Swenson, Jan-Hendrik Prinz, Frank Noé, John D. Chodera, and Peter G. Bolhuis. "OpenPathSampling: A flexible, open framework for path sampling simulations. 1. Basics." J. Chem. Theory Comput. 15, 813 (2019). <https://doi.org/10.1021/acs.jctc.8b00626> David W.H. Swenson, Jan-Hendrik Prinz, Frank Noé, John D. Chodera, and Peter G. Bolhuis. "OpenPathSampling: A flexible, open framework for path sampling simulations. 2. Building and Customizing Path Ensembles and Sample Schemes." J. Chem. Theory Comput. 15, 837 (2019). <https://doi.org/10.1021/acs.jctc.8b00627>; py4vasp has description "Python interface to extract data from VASP calculations"; icet has description "icet is a python tool for the construction and sampling of alloy cluster expansions."; icet has citation M. Ångqvist, W. A. Muñoz, J. M. Rahm, E. Fransson, C. Durniak, P. Rozyczko, T. H. Rod, and P. Erhart ICET – A Python Library for Constructing and Sampling Alloy Cluster Expansions Adv. Theory. Sim., 1900015 (2019) doi: 10.1002/adts.201900015; software has definition source "James Malone"; software has definition source "Modified in parts from <https://en.wikipedia.org/wiki/Software>"; software has definition source "Robert Stevens"

No. 21 answer: pycp2k

No. 22 question: What is the documentation of the "MatDB Online"?

No. 22 context: MatDB has description "MatDB is a database application for experimentally measured engineering materials data. It supports open, registered, and restricted access. It presently hosts more than 20,000 unique data sets coming mainly from European and Member State research programmes. It supports web interfaces for entering, browsing, and retrieving data. MatDB is also enabled for innovative services, including data citation and interoperability standards. The data cit

ation service relies on DataCite DOIs. "; MatPortal has description "Ontology repository for materials science."; MatDB Online has documentation <https://publications.jrc.ec.europa.eu/repository/handle/JRC75978>; MatDB Online has citation Austin T, Over H. MATDB ONLINE—A STANDARDS-BASED SYSTEM FOR PRESERVING, MANAGING, AND EXCHANGING ENGINEERING MATERIALS TEST DATA. DATA SCIENCE JOURNAL 11; 2012. JRC75978; MatDB Online has description "MatDB Online facility is a Standards-based system for preserving, managing, and exchanging engineering materials test data."; Organic Materials Database (OMDB) has description "The organic materials database is an open access electronic structure database for 3-dimensional organic crystals, developed and hosted at the Nordic Institute for Theoretical Physics - Nordita.

It provides tools for search queries based on data-mining and machine learning techniques. The universal features provided on our web interface facilitate the design of novel functional organic materials with a wide-range of applications."; Metadata schemes for materials science data has description "Metadata schemes for materials science data in JSON representation as implemented in TTL representation in the application profiles in the version from 01.05.2022 of the research data management platform CoScInE. The schema are actively developed in the SFB1394 with the aim to construct defect phase diagrams in an automated fashion using all data from advanced experimental characterization and computer simulations produced in this project."; Phonon database at Kyoto university has description "It's getting to be a mainstream idea collecting a big set of ab-initio calculations and analyzing them by informatics approach. For scientific purpose, we plan to do it for phonon properties. As an attempt, we have run a set of ab-initio phonon calculations and left the raw data on this static web site. The aim of this web site is just leaving the data to distribute."; Workshop: Simulation workflows in materials modelling has description "As part of the workshop for Simulation Workflows in Materials Modelling we present a series of pyiron tutorials, starting from a basic introduction to Pyiron, followed by a focus session on up-scaling existing workflows and finally the implementation of new simulation codes and general open science activities of the pyiron project."; Materials Platform for Data Science has description "Online materials database (known as PAULING FILE project) with nearly 2 million entries: physical properties, crystal structures, phase diagrams, available via API, ready for modern data-intensive applications. The source of these entries are about 300,000 peer-reviewed publications in materials science, processed during the last 16 years by an international team of PhD editors. The results are presented online with a quick search interface. The basic access is provided for free."; NOMAD Metainfo has documentation <https://nomad-lab.eu/services/metainfo>; NOMAD Metainfo has description "The NOMAD Metainfo stores descriptive and structured information about materials-science data contained in the NOMAD Archive"; Electronic Structure Common Data Format (ESCDF) has description "This project gathers the working documents used to design the specifications of the Electronic Structure Common Data Format (ESCDF), an initiative of the Electronic Structure Library (ESL) Contributors.

"; Materials Data Repository has description "MDR is a data repository to collect and store papers, presentation materials, and related materials data to accumulate and release them in a form suitable for the promotion of materials research and materials informatics. Users can search the documents and the data from information (metadata) such as sample, instrument, method, and from the full text of the deposited data, to browse and download them freely. "

No. 22 answer: <https://publications.jrc.ec.europa.eu/repository/handle/JRC75978>

No. 23 question: What are the types of software licenses?

No. 23 context: has format has type nan; Proprietary has type Proprietary commercial software license; Commercial license has type Proprietary commercial software license; Proprietary academic license has type Proprietary commercial software license; Academic and commercial license has type Proprietary commercial software license; Proprietary commercial license has type Proprietary commercial software license; Proprietary commercial software license has type nan; Proprietary license has type licence; software repository has type nan; <https://compositions.readthedocs.io/en/latest/> has type documentation; <https://rdf.pyscal.org> has type docu

mentation; <https://fairmat-experimental.github.io/nexus-fairmat-proposal/9636feecb79bb32b828b1a9804269573256d7696/index.html> has type documentation; <https://workflow-gallery.github.io/calphy/README.html> has type documentation; <https://github.com/DataAnalyticsEngineering/EQ2PC#readme> has type documentation; [http://feap.berkeley.edu/wiki/index.php?title=FEAP\\_Wiki\\_Main\\_Page](http://feap.berkeley.edu/wiki/index.php?title=FEAP_Wiki_Main_Page) has type documentation; <http://www.paraview.org/documentation/> has type documentation; [https://www.vasp.at/wiki/index.php/The\\_VASP\\_Manual](https://www.vasp.at/wiki/index.php/The_VASP_Manual) has type documentation; <http://www.castep.org/CASTEP/Documentation> has type documentation; <https://www.cpmc.org/wordpress/index.php/documentation/> has type documentation; <https://docs.abinit.org/> has type documentation; <https://www.quantum-espresso.org/documentation/> has type documentation; <http://jdxft.org/Using.html> has type documentation; <https://manual.cp2k.org/#gsc.tab=0> has type documentation; <https://wiki.fysik.dtu.dk/gpaw/documentation/documentation.html> has type documentation; <http://qboxcode.org/doc/html/> has type documentation; <https://departments.icmab.es/leem/siesta/Documentation/index.html> has type documentation; [http://www.openmx-square.org/openmx\\_man3.9/index.html](http://www.openmx-square.org/openmx_man3.9/index.html) has type documentation; <https://elk.sourceforge.io/#documentation> has type documentation; <http://exciting.wikidot.com/tutorials-oxygen> has type documentation; <https://www.flapw.de/MaX-6.0/documentation/userGuideOverview/> has type documentation; [http://www.wien2k.at/reg\\_user/textbooks/usersguide.pdf](http://www.wien2k.at/reg_user/textbooks/usersguide.pdf) has type documentation; <https://docs.lammps.org/Manual.html> has type documentation; <http://imd.itap.physik.uni-stuttgart.de/userguide/imd.html> has type documentation; <https://manual.gromacs.org/current/index.html> has type documentation; <https://docs.fenicsproject.org/> has type documentation; <https://damask.mpie.de/documentation/index.html> has type documentation; <https://pyiron.readthedocs.io/en/latest/> has type documentation; <https://aiida.readthedocs.io/projects/aiida-core/en/latest/> has type documentation; <http://simstack.readthedocs.io/en/latest/> has type documentation; <https://wiki.fysik.dtu.dk/ase/ase/ase.html> has type documentation; <https://www.ovito.org/docs/current/> has type documentation; <http://jmol.sourceforge.net/docs/> has type documentation; <http://jp-minerals.org/vesta/en/doc.html> has type documentation; <http://www.xcrysden.org/Documentation.html> has type documentation; <https://www.ks.uiuc.edu/Research/vmd/current/ug/> has type documentation; <https://phonopy.github.io/phonopy/#documentation> has type documentation; [https://vibes-developers.gitlab.io/vibes/Documentation/0\\_intro/](https://vibes-developers.gitlab.io/vibes/Documentation/0_intro/) has type documentation; <https://avogadro.cc/docs/> has type documentation; <https://pymatgen.org/usage.html> has type documentation; <http://molmod.github.io/tamkin/> has type documentation; <https://lammps.github.io/pizza/doc/Manual.html> has type documentation; <https://freud.readthedocs.io/en/latest/> has type documentation; <http://molmod.github.io/yaff/> has type documentation; [https://computationalmaterials.gitlab.io/cleese/api\\_doc.html](https://computationalmaterials.gitlab.io/cleese/api_doc.html) has type documentation; <https://kliff.readthedocs.io/en/latest/> has type documentation; <https://mendeleev.readthedocs.io/en/stable/quick.html> has type documentation; <https://ahartmaier.github.io/pyLabFEA/> has type documentation; <https://pacemaker.readthedocs.io/en/latest/> has type documentation; <https://paraprobe-toolbox.readthedocs.io/en/latest/> has type documentation; <https://stabix.readthedocs.io/en/latest/> has type documentation; <https://math.lbl.gov/voro++/doc/> has type documentation; <http://www.qhull.org/html/index.htm> has type documentation; <https://doc.elabftw.net/> has type documentation; <https://kadi4mat.readthedocs.io/en/stable/index.html> has type documentation; <https://github.com/Materials-Data-Science-and-Informatics/metador/blob/main/README.md> has type documentation; <https://docs.httk.org/en/latest/> has type documentation; <https://colabfit.github.io/colabfit-tools/html/index.html> has type documentation; <https://static.oqmd.org/static/docs/index.html> has type documentation; <https://api.materialsproject.org/docs> has type documentation; <https://www.turbomole.org/turbomole/turbomole-documentation/> has type documentation; <http://jacksund.github.io/simmate/home/> has type documentation; <https://kit-data-manager.github.io/fairdoscope/> has type documentation; <https://doi.org/10.1016/j.msea.2019.138295> has type documentation; <https://doi.org/10.1098/rspa.2020.0568> has type documentation; <https://doi.org/10.1002/nme.6869> has type documentation; <http://doi.org/10.3389/fmats.2019.00075> has type documentation; <https://doi.org/10.1007/s10853-020-05740-x> has type documentation; <https://doi.org/10.1016/j.msea.2018.01.033> has type documentation; <https://www.optimade.org/> has type documentation; <https://materials-data-science-and-informatics.github.io/dislocation-ontology/>

has type documentation; w3id.org/mdo/full/1.0/ has type documentation; http://www.quantum-simulation.org/ has type documentation; https://publications.jrc.ec.europa.eu/repository/handle/JRC75978 has type documentation; http://docs.oasis-open.org/materials/materials-matml-spec-pr-01.htm has type documentation; https://www.iucr.org/resources/cif has type documentation; https://colabfit.org/about/ has type documentation; https://citrineinformatics.github.io/gemd-docs/ has type documentation; https://nomad-lab.eu/services/metainfo has type documentation; https://citrineinformatics.github.io/pif-documentation/index.html has type documentation; https://www.nexusformat.org/ has type documentation; https://emmc.eu/moda/ has type documentation; http://icatproject-contrib.github.io/CSMD/ has type documentation; https://www.izus.uni-stuttgart.de/fokus/engmeta has type documentation; https://docs.dftk.org/stable/ has type documentation; https://doi.org/10.5281/zenodo.7625259 has type documentation; https://calphy.org has type documentation; https://pyscal.org has type documentation; https://imagej.net/learn/ has type documentation; documentation has type nan; Commercial license has type Proprietary commercial software license

No. 23 answer: Proprietary commercial software license

No. 24 question: What are the software used to produce the data in the Materials Cloud repository?

No. 24 context: OpenBIS has description "The openBIS platform has three primary functionalities: 1. Inventory management of laboratory samples, materials, protocols, equipment. 2. Laboratory notebook, to document lab experiments. 3. Data management, to store all data related to lab experiments (raw, processed, analysed data, scripts, Jupyter notebooks, etc.)."; Materials Commons has description "A site for Materials Scientists to collaborate, store and publish research. The Materials Commons is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering."; Materials Cloud has conforms to specification OPTIMADE; Materials Cloud Youtube channel required tool Materials Cloud; Materials Cloud has description "Materials Cloud is built to enable the seamless sharing and dissemination of resources in computational materials science, offering educational, research, and archiving tools; simulation software and services; and curated and raw data. These underpin published results and empower data-based discovery, compliant with data management plans and the FAIR principles."; Khazana has description "A platform to store structure and property data created by atomistic simulations, and tool to design materials by learning from the data"; Computational Materials Repository (CMR) has description "Each CMR project consists of an ASE-database and a project page describing the data and showing examples of how to work with the data using Python and ASE."; Semantic Web Technologies (open HPI Massive Open Online Course (MOOC)) has description "In this course, you will learn the fundamentals of Semantic Web technologies and how they are applied for knowledge representation in the World Wide Web. You will learn how to represent knowledge with ontologies and how to access and benefit from semantic data on the Web. Furthermore, you will also learn how to make use of Linked Data and the Web of Data, currently the most popular applications based on Semantic Web technologies."; Semantic Web Technologies (open HPI Massive Open Online Course (MOOC)) has description "In this course, you will learn the fundamentals of Semantic Web technologies. You will learn how to represent knowledge and how to access and benefit from semantic data on the Web."; NOMAD Metainfo has description "The NOMAD Metainfo stores descriptive and structured information about materials-science data contained in the NOMAD Archive"; Materials Data Repository has description "MDR is a data repository to collect and store papers, presentation materials, and related materials data to accumulate and release them in a form suitable for the promotion of materials research and materials informatics. Users can search the documents and the data from information (metadata) such as sample, instrument, method, and from the full text of the deposited data, to browse and download them freely. "; The MatWerk ontology has media type or extend nan; The MatWerk ontology has media type or extend nan; The MatWerk ontology has media type or extend nan; The MatWerk ontology has media type or extend nan; The MatWerk ontology has media type or extend nan; The MatWerk ontology has description "The MatWerk ontology represents research d

ata and related activities of the MSE community. A first version of the ontology will be simplified, focusing on (i) community structure: researchers, research projects, universities, and institutions; (ii) infrastructure: software, workflows, ontologies, schemas, APIs, instruments, facilities, educational resources; and (iii) data: repositories, databases, scientific publications, published datasets and reference data."

No. 24 answer: Materials Cloud

No. 25 question: What are datasets produced by the BAM organization?

No. 25 context: The Materials Data Facility (MDF) has description "A simple way to publish, discover, and access materials datasets"; Computational Materials Repository (CMR) has related resource ASE; Computational Materials Repository (CMR) has description "Each CMR project consists of an ASE-database and a project page describing the data and showing examples of how to work with the data using Python and ASE."; NOMAD MetaInfo has description "The NOMAD MetaInfo stores descriptive and structured information about materials-science data contained in the NOMAD Archive"; Electronic Structure Common Data Format (ESCDF) has repository <https://gitlab.com/ElectronicStructureLibrary/escdf/escdf-specifications>; Electronic Structure Common Data Format (ESCDF) has description "This project gathers the working documents used to design the specifications of the Electronic Structure Common Data Format (ESCDF), an initiative of the Electronic Structure Library (ESL) Contributors.

"; Framework for curation and distribution of reference datasets related participant project is Bundesanstalt für Materialforschung und -prüfung (BAM); Ontologies for defects in crystals related participant project is Bundesanstalt für Materialforschung und -prüfung (BAM)

No. 25 answer: Framework for curation and distribution of reference datasets

No. 26 question: What are some available datasets of mechanical properties of steels?

No. 26 context: Diffusion Database (Kakusan) has description "The Diffusion Database aims to cover all the basic diffusion data of metallic and inorganic materials mainly from the literature sources. The database consists of "diffusion data" and "reference data". Currently the system primarily contains information of pure metals, alloys, semiconductors, ceramics and intermetallics."; Elastic Constant Demo has related resource Elastic Constant Demo Data; Elastic Constant Demo Data has related resource Elastic Constant Demo; Elastic Constant Demo Data has repository <http://hdl.handle.net/21.11102/7f78d0ed-4855-4fdf-b764-0bea725f8821>; Elastic Constant Demo Data has description "Data containing elastic constants of Al alloy"; Inorganic Material Database (AtomWork) has description "The Inorganic Material Database aims to cover all basic crystal structure, x-ray diffraction, property and phase diagram data of inorganic and metallic materials from main literature sources."; Elastic Constant Demo Data has related resource Elastic Constant Demo; Elastic Constant Demo has related resource Elastic Constant Demo Data; Elastic Constant Demo has repository <https://gitlab.mpcdf.mpg.de/smenon/elastic-constant-demo>; Elastic Constant Demo has description "Workflow to calculate elastic constants of Al alloy"; MatDat has description "Over 1000 detailed, fully referenced and verified datasets for steels, aluminium and titanium alloys, cast irons/steels, weld metals. Materials can be searched according to a number of different criteria. Initial search results are presented in the form of a table from which they can be selected for presentation in form of detailed report or for comparison overview (up to 5 materials). In addition to material information and values of properties/parameters, images of microstructure, specimens and those of stress-strain, stress- and strain-life curves (if available) can be reviewed as well."; Atomic Simulation Recipes has related resource ASE; Atomic Simulation Recipes has repository <https://gitlab.com/dtorel/asr>; Atomic Simulation Recipes has description "Recipes for Atomic Scale Materials Research. Collection of python recipes for common (and not so common)

tasks performed in atomic scale materials research. These tasks include relaxation of structures, calculating ground states, calculating band structures, calculating dielectric functions and so on."; Polymer Database (PoLyInfo) has description "Polymer Database "PoLyInfo" systematically provides various

data required for polymeric material design. The main data source is academic literature on polymers. Information on polymers including properties, chemical structures, IUPAC names, processing methods of measured samples, measurement conditions, used monomers and polymerization methods are stored in a object database. About 100 types of properties including thermal, electrical and mechanical properties are covered. Homopolymers, copolymers, furthermore polymer blends, composites and compounds that consist of homopolymers and copolymers are open to the public."

No. 26 answer: MatDat

No. 27 question: What are datasets related to "Transmission electron microscopy"?

No. 27 context: Material Properties Open Database (MPOD) has description "The Material Properties Open Database (MPOD) is a web-based, open access repository of quantitative information about the physical properties of crystalline materials."

MPOD is oriented to design engineers, scientists, science teachers and students. Properties are generally treated as tensor magnitudes. In MPOD the compact matrix notation is applied. To bring an intuitive view of tensor properties, so-called longitudinal properties surfaces are displayed."; pySTEM has description "Automatic segmentation of Scanning Transmission Electron Microscope (STEM) images with unsupervised machine learning"; Soprano has description "Python library for common tasks in computational NMR crystallography and simulations to generate, manipulate, run calculations on and analyse large data sets of crystal structures, with a particular attention to the output of ab-initio random structure searching"; Thermo-Calc has description "Thermo-Calc, is used by materials scientists and engineers to generate material properties data, gain insights about materials, understand a specific observation, and answer direct questions related to a specific material and/or its processing. Used in conjunction with suitable databases, Thermo-Calc can be used for a wide variety of applications."; pyXIT has description "Volume Reconstruction Software for X-ray computed tomography data: reconstruction of set of 2D projections to a 3D volume based on mathematical algorithms"; Atomic Simulation Recipes has related resource ASE; Atomic Simulation Recipes has description "Recipes for Atomic Scale Materials Research. Collection of python recipes for common (and not so common)

tasks performed in atomic scale materials research. These tasks include relaxation of structures, calculating ground states, calculating band structures, calculating dielectric functions and so on."; pycal\_rdf has related resource Computational Material Sample Ontology; Computational Material Sample Ontology has related resource pycal\_rdf; Computational Material Sample Ontology has related Project NFDI-MatWerk; Computational Material Sample Ontology has description "CMSO is an ontology that aims to describe computational materials science samples (or structures), including crystalline defects."

No. 27 answer: pySTEM

No. 28 question: What is the license of the dataset "Elastic Constant Demo Data"?

No. 28 context: NeXus Data Format has license LGPL-2.1 license; NeXus Data Format has documentation <https://www.nexusformat.org/>; NeXus Data Format has description "NeXus is a common data format for neutron, x-ray, and muon science. It is being developed as an international standard by scientists and programmers representing major scientific facilities in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data."; Elastic Constant Demo Data has license BSD 3-Clause License; Elastic Constant Demo Data has file Extension ".h5"; Elastic Constant Demo Data has description "Data containing elastic constants of Al alloy"; Elastic Constant Demo has license BSD 3-Clause License; Elastic Constant Demo has description "Workflow to calculate elastic constants of Al alloy"; TEMMETA has license GNU General Public License v3; TEMMETA has description "TEMMETA is a library for transmission electron microscopy (TEM) (meta)data manipulation. The aim is to offer a one stop place for very basic to intermediate level operations on (S)TEM data, and be a kind of python version of ImageJ + Digital Micrograph + Velox."; Metadata schemes for materials science data has license Creative Commons Attribution 4.0 International; Metadata schemes for materials science data has description "Metadata schemes for materials science data in JSON representation as implemented in TTL representation in the application profiles in



the version from 01.05.2022 of the research data management platform CoScInE. The schema are actively developed in the SFB1394 with the aim to construct defect phase diagrams in an automated fashion using all data from advanced experimental characterization and computer simulations produced in this project."; Phonon database at Kyoto university has license Creative Commons Attribution 4.0 International License; Phonon database at Kyoto university has description "It's getting to be a mainstream idea collecting a big set of ab-initio calculations and analyzing them by informatics approach. For scientific purpose, we plan to do it for phonon properties. As an attempt, we have run a set of ab-initio phonon calculations and left the raw data on this static web site. The aim of this web site is just leaving the data to distribute."; Electronic Structure Common Data Format (ESCDF) has description "This project gathers the working documents used to design the specifications

of the Electronic Structure Common Data Format (ESCDF), an initiative of the Electronic Structure Library (ESL) Contributors.

"; Fiji has license GPLv3+; Fiji has documentation <https://imagej.net/learn/>; Fiji has description "Fiji is an image processing package—a “batteries-included” distribution of ImageJ2, bundling a lot of plugins which facilitate scientific image analysis."

No. 28 answer: BSD 3-Clause License

No. 29 question: What is the repository for "BAM reference data"?

No. 29 context: BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has repository <https://doi.org/10.5281/zenodo.5106606>; BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has file Extension "csv"; BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has file Extension "pdf"; BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has description "The creep behavior of a certified reference material was determined according to ASTM E139-11 using calibrated equipment of an accredited testing laboratory"; Carolina Materials Database has description "Welcome to Carolina Materials Database created in Jianjun Hu's and Ming Hu's groups at University of South Carolina. A freely, globally accessible database of 38,716 inorganic material compounds with over 80,000 calculated properties, and growing."; NOMAD Metainfo has documentation <https://nomad-lab.eu/services/metainfo>; NOMAD Metainfo has description "The NOMAD Metainfo stores descriptive and structured information about materials-science data contained in the NOMAD Archive"

No. 29 answer: <https://doi.org/10.5281/zenodo.5106606>

No. 30 question: What are the different data formats in the "BAM reference data"?

No. 30 context: BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has repository <https://doi.org/10.5281/zenodo.5106606>; BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has file Extension "csv"; BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has file Extension "pdf"; BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has description "The creep behavior of a certified reference material was determined according to ASTM E139-11 using calibrated equipment of an accredited testing laboratory"; Carolina Materials Database has description "Welcome to Carolina Materials Database created in Jianjun Hu's and Ming Hu's groups at University of South Carolina. A freely, globally accessible database of 38,716 inorganic material compounds with over 80,000 calculated properties, and growing."; Computational Materials Repository (CMR) has related resource ASE; Computational Materials Repository (CMR) has description "Each CMR project consists of an ASE-database and a project page describing the data and showing examples of how to work with the data using Python and ASE."; NOMAD Metainfo has documentation <https://nomad-lab.eu/services/metainfo>; NOMAD Metainfo has description "The NOMAD Metainfo stores descriptive and structured information about materials-science data contained in the NOMAD Archive"; Electronic Structure Common Data Format (ESCDF) has repository <https://gitlab.com/ElectronicStructureLibrary/escdf/escdf-specifications>; Elect

ronic Structure Common Data Format (ESCDF) has description "This project gathers the working documents used to design the specifications of the Electronic Structure Common Data Format (ESCDF), an initiative of the Electronic Structure Library (ESL) Contributors."  
"

No. 30 answer: csv; pdf

No. 31 question: What is the software version of "pacemaker"?

No. 31 context: Graphical Expression of Materials Data (GEMD) has description "GEMD stands for Graphical Expression of Materials Data. It's an open-source format initially developed by Citrine Informatics. A version of it is used in their AI platform for materials and chemicals, but as described here it can be used independently."

"; pacemaker has Version "0.2.7"; pacemaker has description "a tool for fitting of interatomic potentials in a general nonlinear Atomic Cluster Expansion (ACE) form."; eLabFTW has Version "4.3.9"; eLabFTW has description "eLabFTW is an electronic lab notebook manager for research teams. It lets you store and organize your research experiments easily. It also features a database where any kind of objects (such as antibodies, plasmids, cell lines, boxes, etc.) can be stored. It is accessed via the browser"; Phonon database at Kyoto university has description "It's getting to be a mainstream idea collecting a big set of ab-initio calculations and analyzing them by informatics approach. For scientific purpose, we plan to do it for phonon properties. As an attempt, we have run a set of ab-initio phonon calculations and left the raw data on this static web site. The aim of this web site is just leaving the data to distribute."; Pyrho has description "Pyrho is a real-space DFT code written in Python. It is not built to be super-fast or scalable- instead, it is built to be super-readable. There's a tutorial notebook (tutorial.ipynb) to guide you through the process of building Pyrho from scratch (starting with the 1D Schrodinger equation). This notebook is based on interactive lectures given by Prof. Joerg Neugebauer."; Qe-tools has programming Language python; Qe-tools has description "A set of useful tools for Quantum ESPRESSO"

No. 31 answer: 0.2.7

No. 32 question: What is the field of research "BAM reference data"?

No. 32 context: BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has discipline high temperature mechanical behavior; BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has description "The creep behavior of a certified reference material was determined according to ASTM E139-11 using calibrated equipment of an accredited testing laboratory"; Diffusion Database (Kakusan) has description "The Diffusion Database aims to cover all the basic diffusion data of metallic and inorganic materials mainly from the literature sources. The database consists of "diffusion data" and "reference data". Currently the system primarily contains information of pure metals, alloys, semiconductors, ceramics and intermetallics."; Dr. Rainer Stotzka has expertise in Research data infrastructure; Dr. Rossella Aversa has expertise in Research data infrastructure; Carolina Materials Database has description "Welcome to Carolina Materials Database created in Jianjun Hu's and Ming Hu's groups at University of South Carolina. A freely, globally accessible database of 38,716 inorganic material compounds with over 80,000 calculated properties, and growing."; :n80e2298a963740ea9bdf3f1a27d3951b21 has first element educational resource; :n80e2298a963740ea9bdf3f1a27d3951b29 has first element educational resource; :n80e2298a963740ea9bdf3f1a27d3951b33 has first element educational resource; Metadata schemes for materials science data has related Project SFB1394; Metadata schemes for materials science data has related Project Project No. 409476157; Metadata schemes for materials science data has description "Metadata schemes for materials science data in JSON representation as implemented in TTL representation in the application profiles in the version from 01.05.2022 of the research data management platform CoScInE. The schema are actively developed in the SFB1394 with the aim to construct defect phase diagrams in an automated fashion using all data from advanced experimental characterization and computer simulations produced in this project."; Phonon database at

Kyoto university has contact point Togo; Phonon database at Kyoto university has description "It's getting to be a mainstream idea collecting a big set of ab-initio calculations and analyzing them by informatics approach. For scientific purpose, we plan to do it for phonon properties. As an attempt, we have run a set of ab-initio phonon calculations and left the raw data on this static web site. The aim of this web site is just leaving the data to distribute."; Workshop: Simulation workflows in materials modelling has description "As part of the workshop for Simulation Workflows in Materials Modelling we present a series of pyiron tutorials, starting from a basic introduction to Pyiron, followed by a focus session on up-scaling existing workflows and finally the implementation of new simulation codes and general open science activities of the pyiron project."; Materials Platform for Data Science has description "Online materials database (known as PAULING FILE project) with nearly 2 million entries: physical properties, crystal structures, phase diagrams, available via API, ready for modern data-intensive applications. The source of these entries are about 300,000 peer-reviewed publications in materials science, processed during the last 16 years by an international team of PhD editors. The results are presented online with a quick search interface. The basic access is provided for free."; :n80e2298a963740ea9bdf3f1a27d3951b6 has first element resource; :n80e2298a963740ea9bdf3f1a27d3951b47 has first element resource; :n80e2298a963740ea9bdf3f1a27d3951b55 has first element resource

No. 32 answer: high temperature mechanical behavior

No. 33 question: What is the description of the "BAM reference data"?

No. 33 context: BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has description "The creep behavior of a certified reference material was determined according to ASTM E139-11 using calibrated equipment of an accredited testing laboratory"; Media Type or Extent has definition "A media type or extent."; Media Type or Extent has definition "The set of conceptual resources specified by the Universal Decimal Classification."; Carolina Materials Database has description "Welcome to Carolina Materials Database created in Jianjun Hu's and Ming Hu's groups at University of South Carolina. A freely, globally accessible database of 38,716 inorganic material compounds with over 80,000 calculated properties, and growing."; Electronic Structure Common Data Format (ESCDF) has description "This project gathers the working documents used to design the specifications

of the Electronic Structure Common Data Format (ESCDF), an initiative of the Electronic Structure Library (ESL) Contributors.

"; Computational Materials Repository (CMR) has description "Each CMR project consists of an ASE-database and a project page describing the data and showing examples of how to work with the data using Python and ASE."

No. 33 answer: The creep behavior of a certified reference material was determined according to ASTM E139-11 using calibrated equipment of an accredited testing laboratory

No. 34 question: What are the datasets produced in 2022?

No. 34 context: nan

An error occurred while processing the problem: can only concatenate str (not "float") to str

No. 34 answer: Unable to get answer

No. 35 question: Who is the creators of the "BAM reference data"?

No. 35 context: BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has creator Sina Schriever; BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has creator Prof. Birgit Skrotzki; BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy has description "The creep behavior of a certified reference material was determined according to ASTM E139-11 using calibrated equipment of an accredited testing laboratory"; AtomisticTools has contributor Interdisciplinary Centre for Advanced Materials Simulation; AtomisticTools has description "This website provides an interactive interface for the assessment of interatomic potentials presented in OpenKIM, NIST IPR and elsewhere with respect to DFT reference data. The calculations were carried out using the pyiron workflow management tool."; Caroli

na Materials Database has description "Welcome to Carolina Materials Database created in Jianjun Hu's and Ming Hu's groups at University of South Carolina. A freely, globally accessible database of 38,716 inorganic material compounds with over 80,000 calculated properties, and growing."; Phonon database at Kyoto university has description "It's getting to be a mainstream idea collecting a big set of ab-initio calculations and analyzing them by informatics approach. For scientific purpose, we plan to do it for phonon properties. As an attempt, we have run a set of ab-initio phonon calculations and left the raw data on this static web site. The aim of this web site is just leaving the data to distribute."; NOMAD MetaInfo has description "The NOMAD MetaInfo stores descriptive and structured information about materials-science data contained in the NOMAD Archive"; Electronic Structure Common Data Format (ESCDF) has description "This project gathers the working documents used to design the specifications

of the Electronic Structure Common Data Format (ESCDF), an initiative of the Electronic Structure Library (ESL) Contributors.

"; The MatWerk ontology has description "The MatWerk ontology represents research data and related activities of the MSE community. A first version of the ontology will be simplified, focusing on (i) community structure: researchers, research projects, universities, and institutions; (ii) infrastructure: software, workflows, ontologies, schemas, APIs, instruments, facilities, educational resources; and (iii) data: repositories, databases, scientific publications, published datasets and reference data."

No. 35 answer: Sina Schrieffer; Prof. Birgit Skrotzki

No. 36 question: What are the datasets published by "BAM"?

No. 36 context: OpenBIS has description "The openBIS platform has three primary functionalities: 1. Inventory management of laboratory samples, materials, protocols, equipment. 2. Laboratory notebook, to document lab experiments. 3. Data management, to store all data related to lab experiments (raw, processed, analysed data, scripts, Jupyter notebooks, etc.)."; Phonon database at Kyoto university has description "It's getting to be a mainstream idea collecting a big set of ab-initio calculations and analyzing them by informatics approach. For scientific purpose, we plan to do it for phonon properties. As an attempt, we have run a set of ab-initio phonon calculations and left the raw data on this static web site. The aim of this web site is just leaving the data to distribute."; Computational Materials Repository (CMR) has related resource ASE; Computational Materials Repository (CMR) has description "Each CMR project consists of an ASE-database and a project page describing the data and showing examples of how to work with the data using Python and ASE."; NOMAD MetaInfo has documentation <https://nomad-lab.eu/services/metainfo>; NOMAD MetaInfo has description "The NOMAD MetaInfo stores descriptive and structured information about materials-science data contained in the NOMAD Archive"; Electronic Structure Common Data Format (ESCDF) has repository <https://gitlab.com/ElectronicStructureLibrary/escdf/escdf-specifications>; Electronic Structure Common Data Format (ESCDF) has description "This project gathers the working documents used to design the specifications

of the Electronic Structure Common Data Format (ESCDF), an initiative of the Electronic Structure Library (ESL) Contributors.

"; Materials Data Repository has description "MDR is a data repository to collect and store papers, presentation materials, and related materials data to accumulate and release them in a form suitable for the promotion of materials research and materials informatics. Users can search the documents and the data from information (metadata) such as sample, instrument, method, and from the full text of the deposited data, to browse and download them freely. "; Framework for curation and distribution of reference datasets related participant project is Bundesanstalt für Materialforschung und -prüfung (BAM); Ontologies for defects in crystals related participant project is Bundesanstalt für Materialforschung und -prüfung (BAM)

No. 36 answer: Bundesanstalt für Materialforschung und -prüfung (BAM)

No. 37 question: Average

No. 37 context: nan

An error occurred while processing the problem: can only concatenate str (not "fl

```
oat") to str
No. 37 answer: Unable to get answer
```

```
In [32]: arr_answers
```

```
Out[32]: ['Prof. Dr. Karsten Durst; Prof. Dr.-Ing. Stephan Wulfinghoff; Christian Dorn;
Dr.-Ing Abril Azocar Guzman; Prof. Dr. Jörg Neugebauer; Prof. Stefan Diebels; P
D Dr. Franz Roters; PD Dr. habil. Thomas Hammerschmidt; Dr. Sarath Menon',
'Project EMMC-CSA (2016-2019); SimDOME (2019-2023); MarketPlace (2018-2022); V
IMMP (2018-2021); OntoTrans (2020-2024); ReaxPro (2019-2023); OntoCommons (2020
-2023); OYSTER (2017-2021); NanoMECommons (2021-2025); OpenModel (2021-2025)',
'Fernández; Prof. Felix Fritzen; Oliver Weeger; Frederik Larsson',
'Mirza Mohtashim Alam',
'support@kitware.com',
'Forschungszentrum Jülich',
'GROMACS; Programming Language: C, C++; Documentation: DOI links provided; Rep
ository: maintained by developers; License: GNU Lesser General Public License',
'Pizza.py Toolkit',
'matminer, Simmate, Pyiron',
'Jan Janssen and colleagues.',
'Interatomic potentials repository (NIST)',
'Pyscal',
'nan',
'dynasor; freud',
'MaterialsMine (MM)\nMaterials Design Ontology (MDO)\nMaterials And Molecules
Basic Ontology (MAMBO)\neNanoMapper ontology',
'A unified multi-physics crystal plasticity simulation package',
'MatPortal',
'Instrument',
'Unable to get answer',
'Infrastructure: software, workflows, ontologies, schemas, APIs, instruments,
facilities, educational resources',
'https://pymatgen.org/',
'pyp2k',
'https://publications.jrc.ec.europa.eu/repository/handle/JRC75978',
'Proprietary commercial software license',
'Materials Cloud',
'Framework for curation and distribution of reference datasets',
'MatDat',
'pySTEM',
'BSD 3-Clause License',
'https://doi.org/10.5281/zenodo.5106606',
'csv; pdf',
'0.2.7',
'high temperature mechanical behavior',
'The creep behavior of a certified reference material was determined according
to ASTM E139-11 using calibrated equipment of an accredited testing laborator
y',
'Unable to get answer',
'Sina Schriever; Prof. Birgit Skrotzki',
'Bundesanstalt für Materialforschung und -prüfung (BAM)',
'Unable to get answer']
```

```
In [33]: df['answer_text'] = arr_answers
df
```

Out[33]:

	Competency Question	Ground Truth	
0	Who is working in the Computational Materials ...	PD Dr. habil. Thomas Hammerschmidt; Prof. Dr. ...	kar
1	What are the research projects associated to E...	VIMMP (2018-2021); OYSTER (2017-2021); SimDOME...	kar
2	Who are the contributors of the data "datasets"?	Prof. Felix Fritzen < <a href="http://demo.fiz-karlsruhe...">http://demo.fiz-karlsruhe...</a>	kar
3	Who is working with Researcher "Ebrahim Norouz..."	Prof. Dr. Harald Sack; Mirza Mohtashim Alam; D...	kar
4	who is the email address of "ParaView"?	support@kitware.com	[('https://schema.org
5	What are the affiliations of Volker Hofmann?	Forschungszentrum Jülich < <a href="http://demo.fiz-karl...">http://demo.fiz-karl...</a>	kar
6	What is "Molecular Dynamics" Software? List th...	1. Resource: <a href="http://demo.fiz-karlsruhe.de/matw...">http://demo.fiz-karlsruhe.de/matw...</a>	kar
7	What are pre- and post-processing tools for MD...	Pizza.py Toolkit; pysical; ASE; MDTraj; freud	kar
8	What are some workflow environments for comput...	Pyiron; AiiDA; SimStack	kar
9	How should I cite pyiron?	"title = {pyiron: An integrated development en...	kar
10	Where can I find a list of interatomic potenti...	Interatomic potentials repository (NIST); Open...	kar
11	What are python libraries used	Pysical	kar

	Competency Question	Ground Truth	
	for calculating...		
12	What are the electronic lab notebooks available?	Pasta ELN; eLabFTW; Karlsruhe Data Infrastruct...	[('https://w3id.org/sk
13	What are the software for Molecular Dynamics (...)	Calphy; Melting temperature computational work...	kar
14	What are the ontologies in nanomaterials domain?	MaterialsMine (MM); NanoParticle Ontology for ...	kar
15	What is DAMASK?	DAMASK is a unified multi-physics crystal plas...	kar
16	What are the data portals for materials scienc...	MatPortal	kar
17	What are the instruments for APT?	No Answer	[('n80e2298a9637
18	In which institution can I find tomography equ...	1. Procon CT-alpha"\n "X-ray micro computed to...	
19	What are the educational resources for Ontology?	Knowledge Graphs (openHPI Massive Open Online ...	kar
20	What is the API of Materials Project?	https://api.materialsproject.org/docs	kar
21	Which simulation software have a python API?	EMMOntoPy	kar
22	What is the documentation of the "MatDB Online"?	https://publications.jrc.ec.europa.eu/reposito...	kar

	Competency Question	Ground Truth	
23	What are the types of software licenses?	BSD-3-Clause license; GPL 3 License; ASL/ BSD-...	
24	What are the software used to produce the data...	pymatgen; AiiDA; Quantum Espresso	kar
25	What are datasets produced by the BAM organiza...	BAM reference data: results of ASTM E139 -11 c...	kar
26	What are some available datasets of mechanical...	Full dataset of several mechanical tests on an...	kar
27	What are datasets related to "Transmission ele...	Radii of S-phase Al <sub>2</sub> CuMg in Al-alloy EN AW-261...	kar
28	What is the license of the dataset "Elastic Co...	BSD 3-Clause License	kar
29	What is the repository for "BAM reference data"?	<a href="http://demo.fiz-karlsruhe.de/matwerk/E291704/">http://demo.fiz-karlsruhe.de/matwerk/E291704;</a> ...	kar
30	What are the different data formats in the "BA...	csv; pdf	kar
31	What is the software version of "pacemaker"?	0.2.7	kar
32	What is the field of research "BAM reference d...	high temperature mechanical behavior	kar
33	What is the description of the "BAM reference ...	The creep behavior of a certified reference ma...	kar
34	What are the datasets	Brinell-Hardness (HBW 2.5/62.5) of Al-alloy EN...	



	Competency Question	Ground Truth
	produced in 2022?	
35	Who is the creators of the "BAM reference data"?	Sina Schriever ( <a href="http://demo.fiz-karlsruhe.de/m...">http://demo.fiz-karlsruhe.de/m...</a> )
36	What are the datasets published by "BAM"?	Datasets for the analysis of dislocations at g...
37	Average	NaN

## Dataframe saved as a XLSX file

```
In [37]: save_path = r"C:\Users\Li\Desktop\results_finetuned.xlsx"

df.to_excel(save_path, index=False)

print("successfully saved", save_path)
```

successfully saved C:\Users\Li\Desktop\results\_finetuned.xlsx

## Evaluation

by Haoran Yang

```
In [38]: import pandas as pd
from rouge import Rouge
from nltk.translate.bleu_score import sentence_bleu
from sentence_transformers import SentenceTransformer, util

# Define paths
excel_file_path = save_path
output_file_path = r"C:\Users\Li\Desktop\results_finetuned_score.xlsx"

# Read Excel file
df = pd.read_excel(excel_file_path, usecols=['Competency Question', 'Context', ''])

# Initialize scorers
rouge = Rouge()
sbert_model = SentenceTransformer('all-MiniLM-L6-v2')

# Prepare a dictionary to store scores
scores_dict = {
    "ROUGE-1-f": [], "ROUGE-1-p": [], "ROUGE-1-r": [],
    "ROUGE-2-f": [], "ROUGE-2-p": [], "ROUGE-2-r": [],
    "ROUGE-L-f": [], "ROUGE-L-p": [], "ROUGE-L-r": [],
```

```

        "BLEU": [],
        "SBERT-Similarity": []
    }

    # Calculate scores
    for _, row in df.iterrows():
        ground_truth = str(row['Ground Truth']) if not pd.isnull(row['Ground Truth']) else ''
        predicted_answer = str(row['answer_text']) if not pd.isnull(row['answer_text']) else ''

        # Calculate ROUGE scores
        if ground_truth and predicted_answer:
            rouge_scores = rouge.get_scores(predicted_answer, ground_truth)[0]
            for key in scores_dict:
                if key.startswith("ROUGE"):
                    rouge_type, rouge_metric = key.split('-')[1], key.split('-')[2]
                    scores_dict[key].append(rouge_scores['rouge-{}'.format(rouge_type, rouge_metric)])
        else:
            for key in scores_dict:
                if key.startswith("ROUGE"):
                    scores_dict[key].append(None)

        # Calculate BLEU score
        reference = ground_truth.split()
        candidate = predicted_answer.split()
        bleu_score = sentence_bleu([reference], candidate, weights=(0.25, 0.25, 0.25, 0.25))
        scores_dict["BLEU"].append(bleu_score)

        # Calculate SBERT-Similarity score
        if ground_truth and predicted_answer:
            embeddings1 = sbert_model.encode([ground_truth], convert_to_tensor=True)
            embeddings2 = sbert_model.encode([predicted_answer], convert_to_tensor=True)
            cos_sim = util.pytorch_cos_sim(embeddings1, embeddings2)
            scores_dict["SBERT-Similarity"].append(cos_sim.item())
        else:
            scores_dict["SBERT-Similarity"].append(None)

    # Add scores to DataFrame
    for key, value in scores_dict.items():
        df[key] = value

    # Save to Excel file
    df.to_excel(output_file_path, index=False)

    print(f"Scores have been calculated and the results are saved to: {output_file_path}")

```

```

c:\users\li\pycharmprojects\pythonproject\venv\lib\site-packages\nltk\translate\b
leu_score.py:552: UserWarning:
The hypothesis contains 0 counts of 3-gram overlaps.
Therefore the BLEU score evaluates to 0, independently of
how many N-gram overlaps of lower order it contains.
Consider using lower n-gram order or use SmoothingFunction()
  warnings.warn(_msg)
c:\users\li\pycharmprojects\pythonproject\venv\lib\site-packages\nltk\translate\b
leu_score.py:552: UserWarning:
The hypothesis contains 0 counts of 4-gram overlaps.
Therefore the BLEU score evaluates to 0, independently of
how many N-gram overlaps of lower order it contains.
Consider using lower n-gram order or use SmoothingFunction()
  warnings.warn(_msg)
c:\users\li\pycharmprojects\pythonproject\venv\lib\site-packages\nltk\translate\b
leu_score.py:552: UserWarning:
The hypothesis contains 0 counts of 2-gram overlaps.
Therefore the BLEU score evaluates to 0, independently of
how many N-gram overlaps of lower order it contains.
Consider using lower n-gram order or use SmoothingFunction()
  warnings.warn(_msg)
Scores have been calculated and the results are saved to: C:\Users\Li\Desktop\res
ults_finetuned_score.xlsx

```

```

In [39]: import pandas as pd
import matplotlib.pyplot as plt

# Load Excel file

output_file_path = r"C:\Users\Li\Desktop\results_finetuned_score.xlsx"
df = pd.read_excel(output_file_path)

# Specify columns to compute mean
columns_to_mean = [
    'BLEU', 'SBERT-Similarity',
    'ROUGE-1-f', 'ROUGE-1-p', 'ROUGE-1-r',
    'ROUGE-2-f', 'ROUGE-2-p', 'ROUGE-2-r',
    'ROUGE-L-f', 'ROUGE-L-p', 'ROUGE-L-r'
]

# Compute mean
means = df[columns_to_mean].mean()

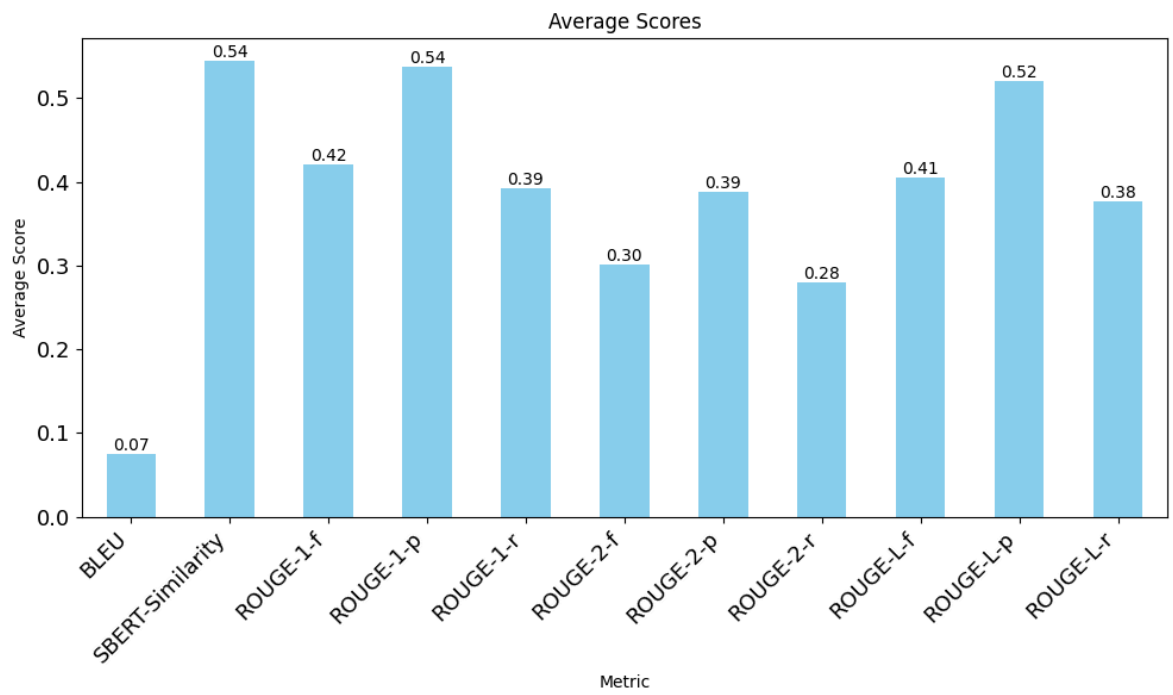
# Visualization
ax = means.plot(kind='bar', figsize=(10, 6), color='skyblue', fontsize=13)

plt.title('Average Scores')
plt.xlabel('Metric')
plt.ylabel('Average Score')
plt.xticks(rotation=45, ha='right') # Rotate x-axis labels for readability
plt.tight_layout() # Automatically adjust subplot parameters to give specified

# Add values above each bar
for i in ax.patches:
    ax.text(i.get_x() + i.get_width() / 2, i.get_height(), f'{i.get_height():.2f}')

plt.show()

```



In [39]:

In [ ]: