



NOVEL - MATERIALS  
DISCOVERY LABORATORY



FRIITZ-HABER-INSTITUT  
MAX-PLANCK-GESELLSCHAFT



# The NOMAD Artificial-Intelligence Toolkit: Web-Based FAIR-Data-Driven Materials Science

[Luigi Sbailò](#)

Humboldt –Universität zu Berlin

# A web based Artificial-Intelligence toolkit

The [NOMAD Archive](#) includes data from [AFLOW](#), [Materials Project](#), [OQMD](#), and more, and from numerous individual researchers.

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, **Materials Project**, **OQMD**, and more, and from numerous individual researchers.

**F**indable **A**ccessible **I**nteroperable **R**eusable (**FAIR**) data

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, **Materials Project**, **OQMD**, and more, and from numerous individual researchers.

➡ Findable **AI Ready** data

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, **Materials Project**, **OQMD**, and more, and from numerous individual researchers.

➡ Findable **AI Ready** data

The *NOMAD Archive* can be queried from the *Artificial-Intelligence Toolkit*, and the data retrieved can be analysed on the fly with the desired AI tools.

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, **Materials Project**, **OQMD**, and more, and from numerous individual researchers.

➔ Findable **AI Ready** data

The *NOMAD Archive* can be queried from the *Artificial-Intelligence Toolkit*, and the data retrieved can be analysed on the fly with the desired AI tools.

The *Artificial-Intelligence Toolkit* hosts tutorials for hands-on learning of the most popular AI tools.

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, **Materials Project**, **OQMD**, and more, and from numerous individual researchers.

➔ Findable **AI Ready** data

The *NOMAD Archive* can be queried from the *Artificial-Intelligence Toolkit*, and the data retrieved can be analysed on the fly with the desired AI tools.

The *Artificial-Intelligence Toolkit* hosts tutorials for hands-on learning of the most popular AI tools.

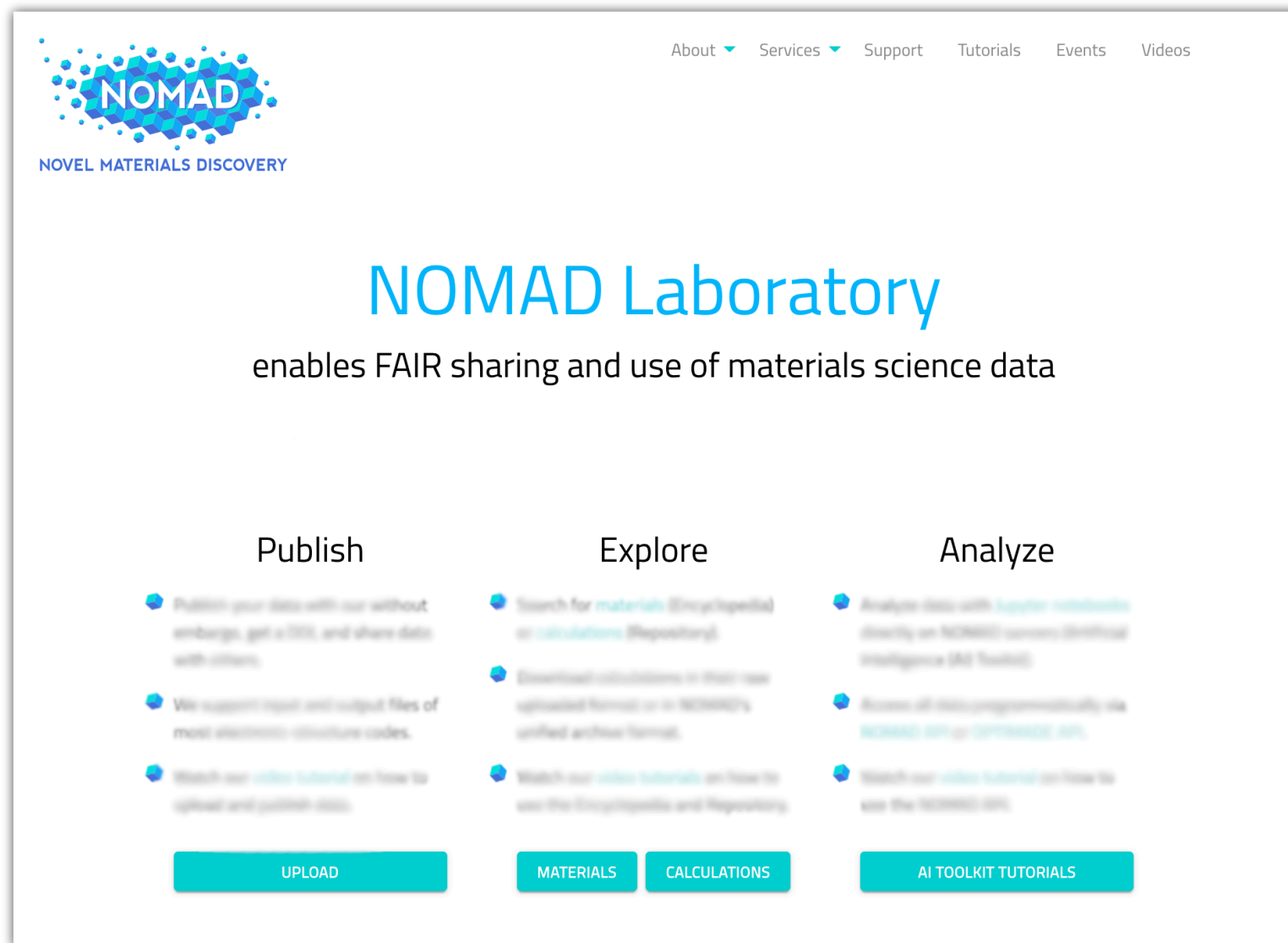
In the *NOMAD Artificial-Intelligence Toolkit*, we maintain several notebooks demonstrating the latest applications of AI to materials science data.

# Technology

- Jupyter Notebook is used as environment for tutorials.
- Machine-learning packages, i.e. scikit-learn, tensorflow, ..., can be imported within the environment for immediate deployment.
- Servers deployed by the NOMAD AI toolkit are hosted by the Max Planck Computer & Data Facility in Garching.







The screenshot shows the homepage of the NOMAD Laboratory. At the top left is the NOMAD logo and the text 'NOVEL MATERIALS DISCOVERY'. At the top right is a navigation menu with links for 'About', 'Services', 'Support', 'Tutorials', 'Events', and 'Videos'. The main heading is 'NOMAD Laboratory' in a large blue font, followed by the tagline 'enables FAIR sharing and use of materials science data'. Below this are three columns of content: 'Publish', 'Explore', and 'Analyze'. Each column contains three bullet points and a teal button at the bottom. The 'Publish' column has a button labeled 'UPLOAD'. The 'Explore' column has two buttons labeled 'MATERIALS' and 'CALCULATIONS'. The 'Analyze' column has a button labeled 'AI TOOLKIT TUTORIALS'.

NOVEL MATERIALS DISCOVERY

About Services Support Tutorials Events Videos

# NOMAD Laboratory

enables FAIR sharing and use of materials science data

## Publish

- Publish your data with us without embargo, get a DOI, and share data with others.
- We support input and output files of most electronic structure codes.
- Watch our video tutorial on how to upload and publish data.

UPLOAD

## Explore

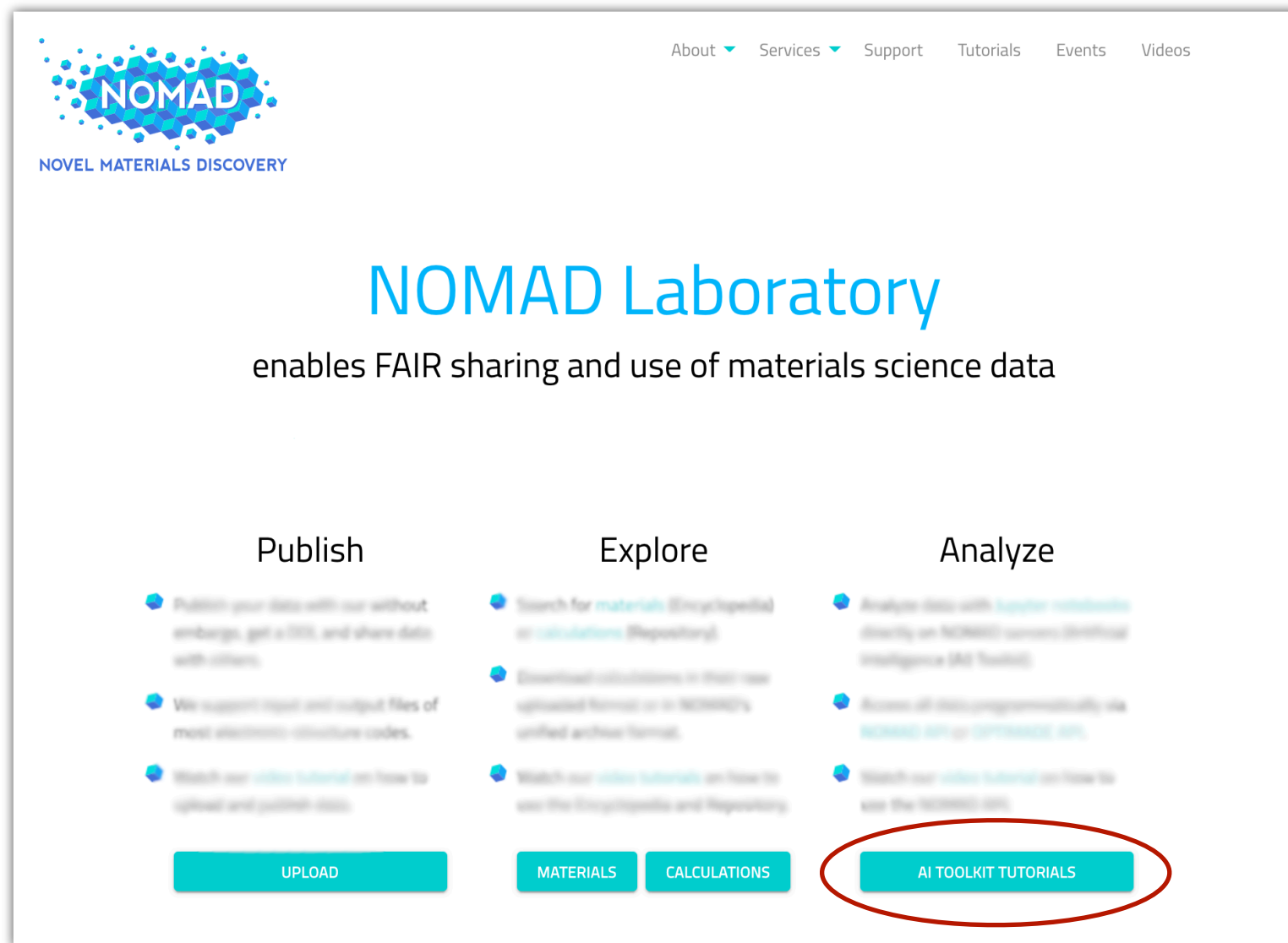
- Search for materials (Encyclopedia) or calculations (Repository).
- Download calculations in their own uploaded format or in NOMAD's unified archive format.
- Watch our video tutorial on how to use the Encyclopedia and Repository.

MATERIALS CALCULATIONS

## Analyze

- Analyze data with transferable models directly on NOMAD's secure Artificial Intelligence (AI Toolkit).
- Access all data programmatically via NOMAD API or OPENMD API.
- Watch our video tutorial on how to use the NOMAD API.

AI TOOLKIT TUTORIALS



The screenshot shows the NOMAD Laboratory website. At the top right, there is a navigation menu with links for 'About', 'Services', 'Support', 'Tutorials', 'Events', and 'Videos'. The main header features the NOMAD logo and the text 'NOVEL MATERIALS DISCOVERY'. The central heading is 'NOMAD Laboratory' in a large blue font, followed by the tagline 'enables FAIR sharing and use of materials science data'. Below this, there are three columns of content: 'Publish', 'Explore', and 'Analyze'. Each column contains a list of bullet points and a button. The 'Analyze' column's button, 'AI TOOLKIT TUTORIALS', is circled in red.

NOVEL MATERIALS DISCOVERY

About Services Support Tutorials Events Videos

# NOMAD Laboratory

enables FAIR sharing and use of materials science data

## Publish

- Publish your data with us without embargo, get a DOI, and share data with others.
- We support input and output files of most electronic structure codes.
- Watch our video tutorial on how to upload and publish data.

UPLOAD

## Explore

- Search for materials (Encyclopedia) or calculations (Repository).
- Download calculations in their own uploaded format or in NOMAD's unified archive format.
- Watch our video tutorial on how to use the Encyclopedia and Repository.

MATERIALS CALCULATIONS

## Analyze

- Analyze data with *Transferable Machine Models* directly on NOMAD's secure *Machine Intelligence (MI Toolkit)*.
- Access all data programmatically via *NOMAD API* or *OPENMDX API*.
- Watch our video tutorial on how to use the NOMAD API.

AI TOOLKIT TUTORIALS

# The NOMAD Artificial-Intelligence toolkit



**Powerful Artificial-Intelligence Tools for Materials Science**

Find New Patterns and Information  
in Materials-Science Big Data

 [Query the archive →](#)

 [View tutorials →](#)

 [Reproduce results →](#)

 [Get to work →](#) 

# The NOMAD Artificial-Intelligence toolkit



**Powerful Artificial-Intelligence Tools for Materials Science**

Find New Patterns and Information in Materials-Science Big Data

Query the archive →

View tutorials →

**Reproduce results →**

Get to work → ⓘ

# Reproducing published results

## ∨ Identifying domains of applicability of machine-Learning models for materials science

Authors: Mohammad-Yasin Arif | Luigi Sbailò | Luca M. Ghiringhelli

## ∧ Discovery of new topological insulators in alloyed tetradymites

Authors: Luigi Sbailò | Thomas A. R. Purcell | Luca M. Ghiringhelli | Matthias Scheffler

Learn how to find descriptive parameters (short formulas) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. This notebook is based on the algorithm 'sure independence screening and sparsifying operator' (SISSO) that enables to search for optimal descriptor by scanning huge feature spaces.

### AI methods:

Supervised learning Classification Symbolic regression Features selection Atomic features SISSO

### System:

Tetradymites Topological insulators

### Additional Resources:

DOI PDF

 ACCESS TUTORIAL

## ∨ ARISE - Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning

Authors: Andreas Leitherer | Angelo Ziletti | Luca M. Ghiringhelli

## ∨ Subgroup discovery of catalysts' genes for carbon-dioxide activation on semiconductor oxides

Authors: Aliaksei Mazheika | Luigi Sbailò | Luca M. Ghiringhelli | Sergey Levchenko | Matthias Scheffler

# Reproducing published results

∨ **Identifying domains of applicability of machine-Learning models for materials science**

Authors: Mohammad-Yasin Arif | Luigi Sbailò | Luca M. Ghiringhelli

∧ **Discovery of new topological insulators in alloyed tetradymites**

Authors: Luigi Sbailò | Thomas A. R. Purcell | Luca M. Ghiringhelli | Matthias Scheffler

Learn how to find descriptive parameters (short formulas) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. This notebook is based on the algorithm 'sure independence screening and sparsifying operator' (SISSO) that enables to search for optimal descriptor by scanning huge feature spaces.

AI methods:

Supervised learning Classification Symbolic regression Features selection Atomic features

Guohua Cao et al., Phys. Rev. Materials 4, 034204 (2020)

 ACCESS TUTORIAL

Additional Resources:

[DOI](#) [PDF](#)

∨ **ARISE - Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning**

Authors: Andreas Leitherer | Angelo Ziletti | Luca M. Ghiringhelli

∨ **Subgroup discovery of catalysts' genes for carbon-dioxide activation on semiconductor oxides**

Authors: Aliaksei Mazheika | Luigi Sbailò | Luca M. Ghiringhelli | Sergey Levchenko | Matthias Scheffler

# Reproducing published results

## Identifying domains of applicability of machine-Learning models for materials science

Authors: Mohammad-Yasin Arif | Luigi Sbailò | Luca M. Ghiringhelli

## Discovery of new topological insulators in alloyed tetradymites

Authors: Luigi Sbailò | Thomas A. R. Purcell | Luca M. Ghiringhelli | Matthias Scheffler

Learn how to find descriptive parameters (short formulas) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. This notebook is based on the algorithm 'sure independence screening and sparsifying operator' (SISSO) that enables to search for optimal descriptor by scanning huge feature spaces.

### AI methods:

Supervised learning Classification Symbolic regression Features selection Atomic features

Guohua Cao et al., Phys. Rev. Materials 4, 034204 (2020)



### Additional Resources:



## ARISE - Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning

Authors: Andreas Leitherer | Angelo Ziletti | Luca M. Ghiringhelli

## Subgroup discovery of catalysts' genes for carbon-dioxide activation on semiconductor oxides


Authors: Aliaksei Mazheika | Luigi Sbailò | Luca M. Ghiringhelli | Sergey Levchenko | Matthias Scheffler



# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**  
created by: Luigi Sbailò, Thomas A. R. Purcell, Luca M. Ghiringhelli, and Matthias Scheffler

Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany  
Last updated: Sep 29, 2020



**Introduction**  
This tutorial shows how to find descriptor parameters (short formulas) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. It is based on the algorithm suite independence screening and sparsifying operator (SSISO), that enables to search for optimal descriptor by scanning huge feature spaces.

**Reproduction of the method used to reproduce the results**  
The idea demonstrated in this tutorial is to start from simple physical quantities ("simple features"), here properties of the constituent free atoms such as Pauling electronegativity, to provide reference to classes of candidate formulae by applying arithmetic operations (addition, subtraction, multiplication, division, exponentiation, logarithm, square root, etc.) to generate a large number of candidate formulae. These candidate formulae are then screened by the SSISO operator. The SSISO is used to select only a few of these formulae that explain the data.

**Reproduction of the method used to reproduce the results**  
The idea demonstrated in this tutorial is to start from simple physical quantities ("simple features"), here properties of the constituent free atoms such as Pauling electronegativity, to provide reference to classes of candidate formulae by applying arithmetic operations (addition, subtraction, multiplication, division, exponentiation, logarithm, square root, etc.) to generate a large number of candidate formulae. These candidate formulae are then screened by the SSISO operator. The SSISO is used to select only a few of these formulae that explain the data.

## Discovery of new topological insulators in alloyed tetradymites

created by: Luigi Sbailò, Thomas A. R. Purcell, Luca M. Ghiringhelli, and Matthias Scheffler

Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany  
[Last updated: Sep 29, 2020]



MAX-PLANCK-GESELLSCHAFT



### Introduction

This tutorial shows how to find descriptor parameters (short formulas) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. It is based on the algorithm suite independence screening and sparsifying operator (SSISO), that enables to search for optimal descriptor by scanning huge feature spaces.

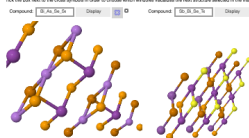
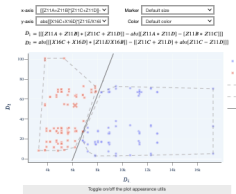
R. Ouyang, S. Curtarolo, E. Arrighetti, M. Scheffler, L. M. Ghiringhelli: SSISO: a compressed sensing method for identifying the best low-dimensional descriptor in an immensity of offshoot candidates, *Phys. Rev. Materials* 2, 033602 (2018) [PDF].

With the default settings, the method reproduces the same results from:

G. Cao, R. Ouyang, L. M. Ghiringhelli, M. Scheffler, H. Liu, C. Carbogno, and Z. Zhang: *Artificial intelligence for high-throughput discovery of topological insulators: The example of alloyed tetradymites*, *Phys. Rev. Materials* 4, 034204 (2020) [PDF].

Options	Feature	Settings	Default selection
<input type="checkbox"/> $x+y$	$x$	1000000	<input type="checkbox"/>
<input type="checkbox"/> $x-y$	$x$	1000000	<input type="checkbox"/>
<input type="checkbox"/> $x \cdot y$	$x$	To optimize the feature weights	<input type="checkbox"/>
<input type="checkbox"/> $x/y$	$x$	Same as for the feature weights	<input type="checkbox"/>
<input type="checkbox"/> $\log(x)$	$x$	Number of features per SSISO iteration	<input type="checkbox"/>
<input type="checkbox"/> $\exp(x)$	$x$	Iteration number of iterations	<input type="checkbox"/>
<input type="checkbox"/> $x^2$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $x^3$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $x^4$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $x^5$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $\log(x^2)$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $\log(x^3)$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $\log(x^4)$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $\log(x^5)$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $\log(x^6)$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $\log(x^7)$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $\log(x^8)$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $\log(x^9)$	$x$		<input type="checkbox"/>
<input type="checkbox"/> $\log(x^{10})$	$x$		<input type="checkbox"/>

Number of features generated: 10000  
SSISO results:  
Selected features:  $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{36}, x_{37}, x_{38}, x_{39}, x_{40}, x_{41}, x_{42}, x_{43}, x_{44}, x_{45}, x_{46}, x_{47}, x_{48}, x_{49}, x_{50}, x_{51}, x_{52}, x_{53}, x_{54}, x_{55}, x_{56}, x_{57}, x_{58}, x_{59}, x_{60}, x_{61}, x_{62}, x_{63}, x_{64}, x_{65}, x_{66}, x_{67}, x_{68}, x_{69}, x_{70}, x_{71}, x_{72}, x_{73}, x_{74}, x_{75}, x_{76}, x_{77}, x_{78}, x_{79}, x_{80}, x_{81}, x_{82}, x_{83}, x_{84}, x_{85}, x_{86}, x_{87}, x_{88}, x_{89}, x_{90}, x_{91}, x_{92}, x_{93}, x_{94}, x_{95}, x_{96}, x_{97}, x_{98}, x_{99}, x_{100}$





# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**  
 authors: Luigi Sbailò, Thomas A. Friedl, Lucian Comanescu, and Thomas Heine  
 FHO Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany  
 last updated: May 24, 2020

**Introduction**  
 This feature shows how to find descriptor parameters (short forms) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. It is based on the algorithmic independence screening and sparsifying operator (SISSO), that enables to search for optimal descriptors by learning high feature spaces.

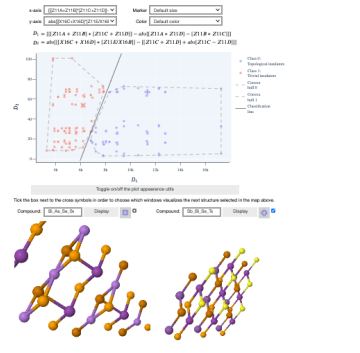
**Number of features generated: 15000**

Operations:	Features:	Settings:	Default selection
<input checked="" type="checkbox"/> $x + y$	<input checked="" type="checkbox"/> $Z_{cations}$	SISSO rung: PRM2020	Run
<input checked="" type="checkbox"/> $x - y$	<input checked="" type="checkbox"/> $\chi_{cations}$	To unfreeze the feature selection, please select any rung other than PRM2020.	Plot interactive map
<input checked="" type="checkbox"/> $ x - y $	<input checked="" type="checkbox"/> $\lambda_{cations}$	Number of selected features per SIS iteration: 50	
<input checked="" type="checkbox"/> $x \cdot y$	<input checked="" type="checkbox"/> $Z_{anions}$	Maximum number of dimensions: 2	
<input checked="" type="checkbox"/> $x/y$	<input checked="" type="checkbox"/> $\chi_{anions}$		
<input checked="" type="checkbox"/> $\exp(x)$	<input checked="" type="checkbox"/> $\lambda_{anions}$		
<input checked="" type="checkbox"/> $\exp(-x)$			
<input checked="" type="checkbox"/> $1/x$			
<input checked="" type="checkbox"/> $x^2$			
<input checked="" type="checkbox"/> $x^3$			
<input checked="" type="checkbox"/> $\sqrt{x}$			
<input checked="" type="checkbox"/> $\sqrt[3]{x}$			
<input checked="" type="checkbox"/> $\log(x)$			
<input checked="" type="checkbox"/> $ x $			

Number of features generated: 15000

1D model  
 # misclassified: 25  
 SVM dividing line:  $c_0 + a_0 * \frac{[Z11A*Z11B]}{\sqrt{[X16E]}} / ([Z11A/Z11C] + [Z11B/Z11D]) = 0$   
 $c_0: 8.968 \quad | \quad a_0: -0.01492$

2D model  
 # misclassified: 0  
 SVM dividing line:  $c_0 + a_0 * \text{abs}([X16C+X16D] * [Z11E/X16B]) - [Z11C+Z11D] + \text{abs}[Z11C-Z11D] + a_1 * ([Z11A+Z11B] * [Z11C+Z11D]) - \text{abs}([Z11A*Z11D] - [Z11B*Z11C]) = 0$   
 $c_0: 25.55 \quad | \quad a_0: 0.1313 \quad | \quad a_1: -0.004255$



# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**  
 authors: Luigi Sbailò, Thomas A. Friedl, Ludvig Öngren, and Thomas Gessner  
 FHO Hohen Heide of the Max Planck Society, Farnbergweg 4, D-14183 Berlin, Germany  
 last updated: 2022-04-28

**Introduction**  
 This feature shows how to find descriptor parameters (short forms) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. It is based on the algorithm sure independence screening and sparsifying operator (SISSO), that enables to search for optimal descriptors for learning high feature spaces.

**Number of features generated: 15000**

Operations:	Features:	Settings:
<input checked="" type="checkbox"/> $x + y$	<input checked="" type="checkbox"/> $Z_{cations}$	SISSO rung: PRM2020
<input checked="" type="checkbox"/> $x - y$	<input checked="" type="checkbox"/> $\chi_{cations}$	To unfreeze the feature selection, please select any rung other than PRM2020.
<input checked="" type="checkbox"/> $ x - y $	<input checked="" type="checkbox"/> $\lambda_{cations}$	Number of selected features per SIS iteration: 50
<input checked="" type="checkbox"/> $x \cdot y$	<input checked="" type="checkbox"/> $Z_{anions}$	Maximum number of dimensions: 2
<input checked="" type="checkbox"/> $x/y$	<input checked="" type="checkbox"/> $\chi_{anions}$	
<input checked="" type="checkbox"/> $\exp(x)$	<input checked="" type="checkbox"/> $\lambda_{anions}$	
<input checked="" type="checkbox"/> $\exp(-x)$		
<input checked="" type="checkbox"/> $1/x$		
<input checked="" type="checkbox"/> $x^2$		
<input checked="" type="checkbox"/> $x^3$		
<input checked="" type="checkbox"/> $\sqrt{x}$		
<input checked="" type="checkbox"/> $\sqrt[3]{x}$		
<input checked="" type="checkbox"/> $\log(x)$		
<input checked="" type="checkbox"/> $ x $		

Default selection

Run

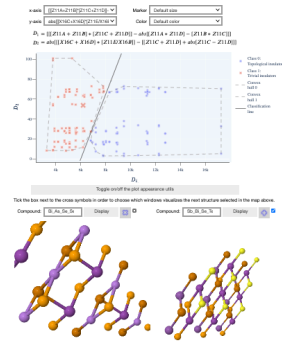
Plot interactive map

## Input parameters

Number of features generated: 15000

1D model  
 # misclassified: 25  
 SVM dividing line:  $c_0 + a_0 * \frac{[Z11A*Z11B]}{\sqrt{[X16E]}} / ([Z11A/Z11C] + [Z11B/Z11D]) = 0$   
 $c_0: 8.968 \quad | \quad a_0: -0.01492$

2D model  
 # misclassified: 0  
 SVM dividing line:  $c_0 + a_0 * \text{abs}([X16C + X16D] * [Z11E/X16B]) - [Z11C + Z11D] + \text{abs}[Z11C - Z11D] + a_1 * ([Z11A + Z11B] * [Z11C + Z11D]) - \text{abs}([Z11A * Z11D] - [Z11B * Z11C]) = 0$   
 $c_0: 25.55 \quad | \quad a_0: 0.1313 \quad | \quad a_1: -0.004255$



# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**  
 authors: Luigi Sbailò, Thomas A. Friedl, Lucian Oprea, and Thomas Heine  
 FHO Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany  
 last updated: May 24, 2020

**Introduction**  
 This article shows how to find descriptor parameters (short forms) that predict whether alloyed materials are topological or trivial materials, using the example of tetradymites. It is based on the algorithmic independence screening and sparsifying operator (SISSO), that enables to search for optimal descriptors by learning high feature spaces.

**Operations:**  
 The idea demonstrated in this article is to start from simple physical quantities ("primary features"), learn properties of the considered free elements such as Pauling electronegativity, to provide reference to detect of candidate features by applying arithmetic operations (addition, subtraction, multiplication, division, exponentiation) to generate a large number of features. These candidate features are then filtered by the SISSO method. The SISSO is used to select only a few of these formulas that explain the data.

**Features:**  
 By choosing descriptors for "primary features", we can understand the SISSO method by the SISSO method. The SISSO method consists of two main steps: feature selection and model selection. In the first step, the SISSO method generates a large number of candidate features by applying arithmetic operations (addition, subtraction, multiplication, division, exponentiation) to the primary features. In the second step, the SISSO method selects the best features by using the SISSO method.

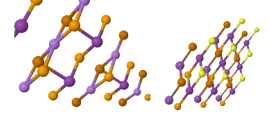
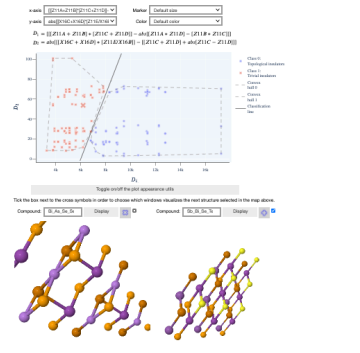
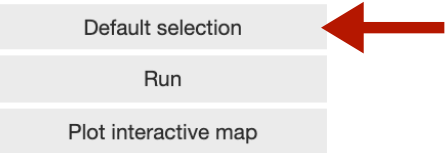
**Settings:**  
 The SISSO method is implemented in the SISSO software. The SISSO software is available at <https://github.com/luigi-sbailo/sisso>. The SISSO software is implemented in the SISSO software. The SISSO software is available at <https://github.com/luigi-sbailo/sisso>.

Operations:	Features:	Settings:
<input checked="" type="checkbox"/> $x + y$	<input checked="" type="checkbox"/> $Z_{cations}$	SISSO rung: PRM2020
<input checked="" type="checkbox"/> $x - y$	<input checked="" type="checkbox"/> $\chi_{cations}$	To unfreeze the feature selection, please select any rung other than PRM2020.
<input checked="" type="checkbox"/> $ x - y $	<input checked="" type="checkbox"/> $\lambda_{cations}$	Number of selected features per SIS iteration: 50
<input checked="" type="checkbox"/> $x \cdot y$	<input checked="" type="checkbox"/> $Z_{anions}$	Maximum number of dimensions: 2
<input checked="" type="checkbox"/> $x/y$	<input checked="" type="checkbox"/> $\chi_{anions}$	
<input checked="" type="checkbox"/> $\exp(x)$	<input checked="" type="checkbox"/> $\lambda_{anions}$	
<input checked="" type="checkbox"/> $\exp(-x)$		
<input checked="" type="checkbox"/> $1/x$		
<input checked="" type="checkbox"/> $x^2$		
<input checked="" type="checkbox"/> $x^3$		
<input checked="" type="checkbox"/> $\sqrt{x}$		
<input checked="" type="checkbox"/> $\sqrt[3]{x}$		
<input checked="" type="checkbox"/> $\log(x)$		
<input checked="" type="checkbox"/> $ x $		

Number of features generated: 15000

1D model  
 # misclassified: 25  
 SVM dividing line:  $c_0 + a_0 * \frac{[Z11A*Z11B]}{\sqrt{[X16E]}} / ([Z11A/Z11C] + [Z11B/Z11D]) = 0$   
 $c_0: 8.968 \quad | \quad a_0: -0.01492$

2D model  
 # misclassified: 0  
 SVM dividing line:  $c_0 + a_0 * \text{abs}([X16C+X16D] * [Z11E/X16B]) - [Z11C+Z11D] + \text{abs}[Z11C-Z11D] + a_1 * ([Z11A+Z11B] * [Z11C+Z11D]) - \text{abs}([Z11A*Z11D] - [Z11B*Z11C]) = 0$   
 $c_0: 25.55 \quad | \quad a_0: 0.1313 \quad | \quad a_1: -0.004255$



# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**  
 authors: Luigi Sbailò, Thomas A. Friedl, Ludovic D'arrigo, and Thomas Frauenheim  
 FHO Hohen Heide of the Max-Planck Society, Faradayweg 4, D-14195 Berlin, Germany  
 last updated: 2024-04-20

**Introduction**  
 This article shows how to find descriptor parameters (short forms) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. It is based on the algorithmic independence screening and sparsifying operator (SISSO), that enables to search for optimal descriptors by learning high feature spaces.

**Number of features generated: 15000**

- |   |   |
|---|---|
| <b>Operations:</b>                                | <b>Features:</b>  |
| <input checked="" type="checkbox"/> $x + y$       | <input checked="" type="checkbox"/> $Z_{cations}$       |
| <input checked="" type="checkbox"/> $x - y$       | <input checked="" type="checkbox"/> $\chi_{cations}$    |
| <input checked="" type="checkbox"/> $ x - y $     | <input checked="" type="checkbox"/> $\lambda_{cations}$ |
| <input checked="" type="checkbox"/> $x \cdot y$   | <input checked="" type="checkbox"/> $Z_{anions}$        |
| <input checked="" type="checkbox"/> $x/y$         | <input checked="" type="checkbox"/> $\chi_{anions}$     |
| <input checked="" type="checkbox"/> $\exp(x)$     | <input checked="" type="checkbox"/> $\lambda_{anions}$  |
| <input checked="" type="checkbox"/> $\exp(-x)$    |   |
| <input checked="" type="checkbox"/> $1/x$         |   |
| <input checked="" type="checkbox"/> $x^2$         |   |
| <input checked="" type="checkbox"/> $x^3$         |   |
| <input checked="" type="checkbox"/> $\sqrt{x}$    |   |
| <input checked="" type="checkbox"/> $\sqrt[3]{x}$ |   |
| <input checked="" type="checkbox"/> $\log(x)$     |   |
| <input checked="" type="checkbox"/> $ x $         |   |

**Settings:**

SISSO rung:

To unfreeze the feature selection, please select any rung other than PRM2020.

Number of selected features per SIS iteration:

Maximum number of dimensions:

Default selection

Run

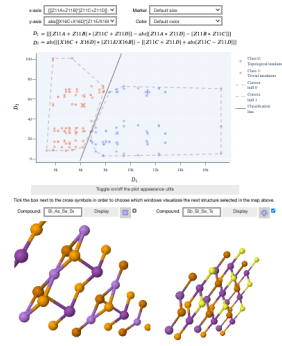
Plot interactive map

Output

Number of features generated: 15000

1D model  
 # misclassified: 25  
 SVM dividing line:  $c_0 + a_0 * \frac{[Z11A * Z11B]}{\sqrt{[X16E]}} / ([Z11A / Z11C] + [Z11B / Z11D]) = 0$   
 $c_0: 8.968 \quad | \quad a_0: -0.01492$

2D model  
 # misclassified: 0  
 SVM dividing line:  $c_0 + a_0 * \text{abs}([X16C + X16D] * [Z11E / X16B]) - [Z11C + Z11D] + \text{abs}[Z11C - Z11D] + a_1 * ([Z11A + Z11B] * [Z11C + Z11D]) - \text{abs}([Z11A * Z11D] - [Z11B * Z11C]) = 0$   
 $c_0: 25.55 \quad | \quad a_0: 0.1313 \quad | \quad a_1: -0.004255$



# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**  
 authors: Luigi Sbailò, Thomas A. Friedl, Lucian Ciomperli, and Thomas Frauenheim  
 FZI-Haber Institute of the Max-Planck Society, Faradayweg 4, D-14195 Berlin, Germany  
 last updated: May 24, 2022

**Introduction**  
 This feature shows how to find descriptor parameters (short forms) that predict whether alloyed materials are topological or trivial materials, using the example of tetradymites. It is based on the algorithmic independent screening and sparsifying operator (SISSO), that enables to search for optimal descriptors by learning high feature spaces.

**Number of features generated: 15000**

- Operations:**
- $x + y$
  - $x - y$
  - $|x - y|$
  - $x \cdot y$
  - $x/y$
  - $\exp(x)$
  - $\exp(-x)$
  - $1/x$
  - $x^2$
  - $x^3$
  - $\sqrt{x}$
  - $\sqrt[3]{x}$
  - $\log(x)$
  - $|x|$

- Features:**
- $Z_{cations}$
  - $\chi_{cations}$
  - $\lambda_{cations}$
  - $Z_{anions}$
  - $\chi_{anions}$
  - $\lambda_{anions}$

**Settings:**

SISSO rung:

To unfreeze the feature selection, please select any rung other than PRM2020.

Number of selected features per SIS iteration:

Maximum number of dimensions:

Number of features generated: 15000

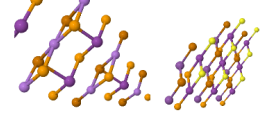
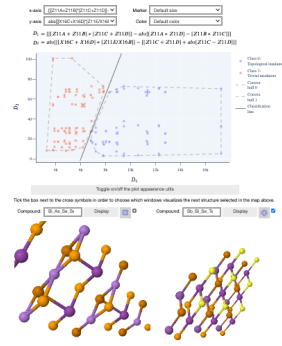
1D model  
 # misclassified: 25  
 SVM dividing line:  $c_0 + a_0 * \frac{[Z11A*Z11B]}{\sqrt{[X16E]}} / ([Z11A/Z11C] + [Z11B/Z11D]) = 0$   
 $c_0: 8.968 \quad | \quad a_0: -0.01492$

2D model  
 # misclassified: 0  
 SVM dividing line:  $c_0 + a_0 * \text{abs}([X16C + X16D] * [Z11E/X16B]) - [Z11C + Z11D] + \text{abs}[Z11C - Z11D] + a_1 * ([Z11A + Z11B] * [Z11C + Z11D]) - \text{abs}([Z11A * Z11D] - [Z11B * Z11C]) = 0$   
 $c_0: 25.55 \quad | \quad a_0: 0.1313 \quad | \quad a_1: -0.004255$

Default selection

Run

Plot interactive map







# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**  
 authors: Luigi Sbailò, Thomas A. Friedl, Lorenz Lindner, and Thomas Chatterjee  
 FHO Höherer Institute of the Max Planck Society, Faradayweg 4, D-14195 Berlin, Germany  
 last updated: May 24, 2020

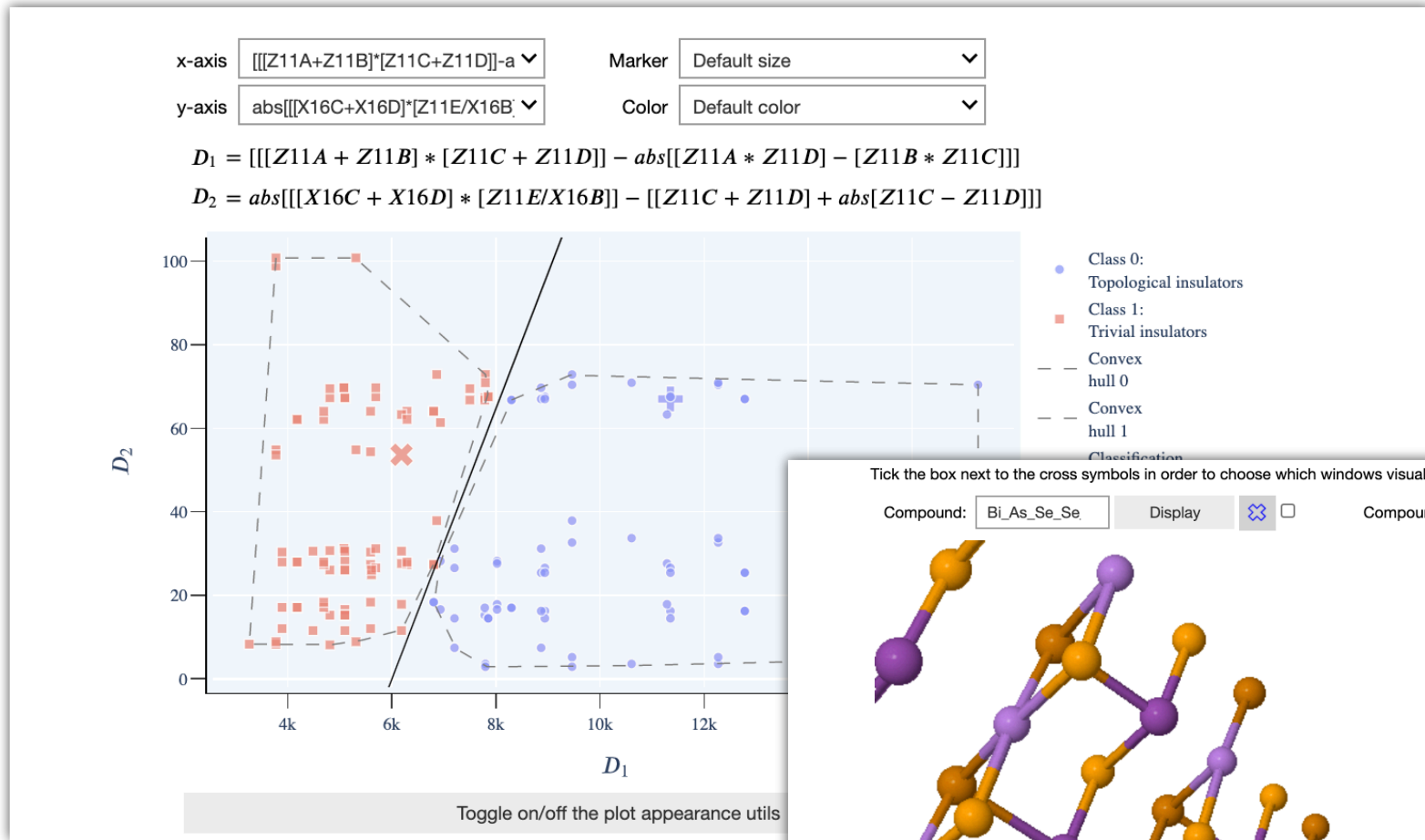
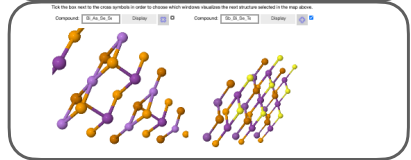
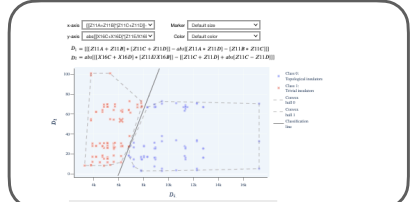
**Introduction**  
 The article shows how to find descriptor parameters (short formulae) that predict whether alloyed tetradymites are topological or trivial insulators, using the example of SbTe<sub>2</sub>Te. It is based on the algorithmic search independence screening and quantifying operator (SSISO), that enables to search for optimal descriptors by learning high feature spaces.

**Features of the method used for reproduction**  
 The idea demonstrated in this article is to start from simple physical quantities ("primary features"), learn properties of the compound free elements such as Pauling electronegativity, to provide reference or baseline of variables formed by applying arithmetic operations (primary features). These variables form the space for the search "feature space". The SSISO is used to select only a few of these formulae that explain the data.

**Implementation of the method used for reproduction**  
 The idea demonstrated in this article is to start from simple physical quantities ("primary features"), learn properties of the compound free elements such as Pauling electronegativity, to provide reference or baseline of variables formed by applying arithmetic operations (primary features). These variables form the space for the search "feature space". The SSISO is used to select only a few of these formulae that explain the data.

Options	Feature	Settings	Default selection
<input type="checkbox"/> A + B	$\chi_{\text{Pauling}}$	SSISO rule	<input type="checkbox"/>
<input type="checkbox"/> A + C	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + D	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + E	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + F	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + G	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + H	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + I	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + J	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + K	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + L	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + M	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + N	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + O	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + P	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + Q	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + R	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + S	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + T	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + U	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + V	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + W	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + X	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + Y	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + Z	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AA	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AB	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AC	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AD	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AE	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AF	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AG	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AH	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AI	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AJ	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AK	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AL	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AM	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AN	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AO	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AP	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AQ	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AR	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AS	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AT	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AU	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AV	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AW	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AX	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AY	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AZ	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AA	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AB	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AC	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AD	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AE	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AF	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AG	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AH	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AI	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AJ	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AK	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AL	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AM	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AN	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AO	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AP	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AQ	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AR	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AS	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AT	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AU	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AV	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AW	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AX	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AY	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AZ	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AA	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AB	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AC	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AD	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AE	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AF	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AG	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AH	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AI	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AJ	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AK	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AL	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AM	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AN	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AO	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AP	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AQ	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AR	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AS	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AT	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AU	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AV	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AW	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AX	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AY	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>
<input type="checkbox"/> A + AZ	$\chi_{\text{Pauling}}$	Pauling electronegativity	<input type="checkbox"/>

Number of features generated: 19100  
 SSISO rule:  $D_1 = \frac{[Z11A + Z11B] * [Z11C + Z11D]}{[X16C + X16D]}$   
 SSISO rule:  $D_2 = \frac{abs([X16C + X16D] * [Z11E/X16B]) - ([Z11C + Z11D] + abs([Z11C - Z11D]))}{[Z11A * Z11D] - [Z11B * Z11C]}$

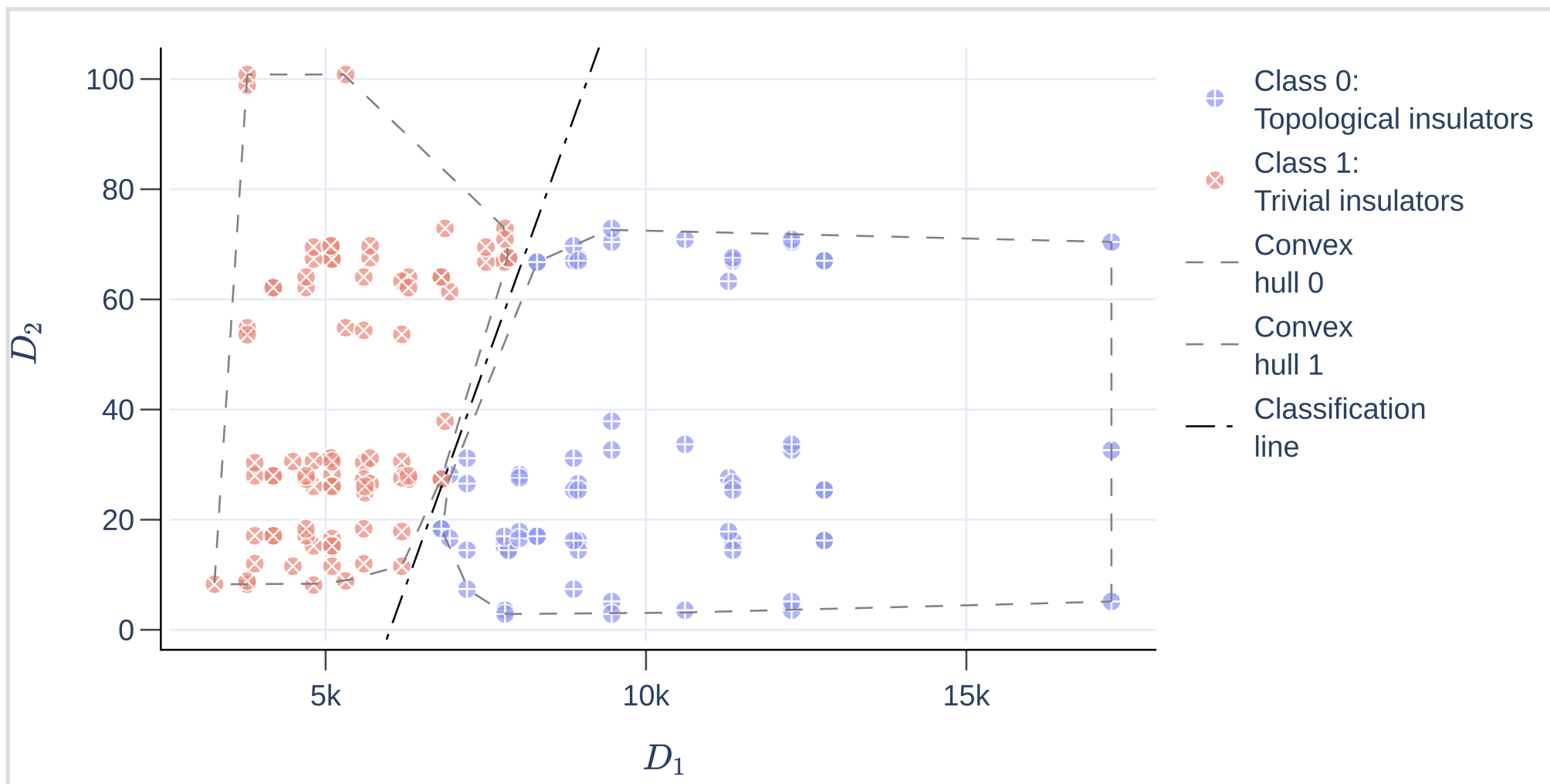


Tick the box next to the cross symbols in order to choose which windows visualizes the next structure selected in the map above.

Compound: Bi<sub>2</sub>As<sub>2</sub>Se<sub>2</sub>Te<sub>2</sub> Display

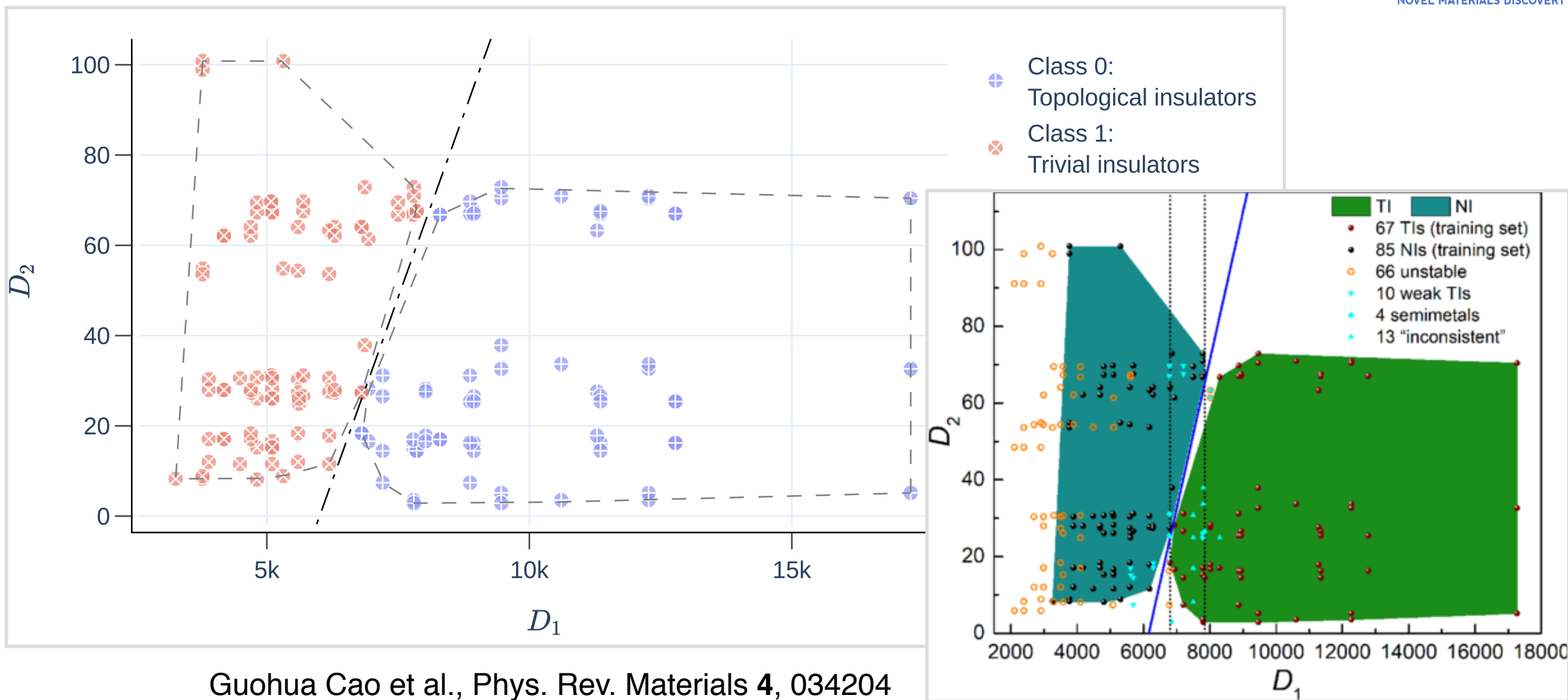
Compound: Sb<sub>2</sub>Bi<sub>2</sub>Se<sub>2</sub>Te<sub>2</sub> Display

# Reproducing published results





# Reproducing published results



## Publish

- Publish your data with our without embargo, get a DOI, and share data with others.
- We support input and output files of most electronic-structure codes.
- Watch our [video tutorial](#) on how to upload and publish data.

UPLOAD

## Explore

- Search for [materials](#) (Encyclopedia) or [calculations](#) (Repository).
- All raw and processed data can be downloaded and used under the [CC BY 4.0](#).
- Watch our [video tutorials](#) on how to use the Encyclopedia and Repository.

MATERIALS

CALCULATIONS

## Analyze

- Analyze data with [Jupyter notebooks](#) directly on NOMAD servers (Artificial Intelligence (AI) Toolkit).
- Access all data programmatically via [NOMAD API](#) or [OPTIMADE API](#).
- Watch our [video tutorial](#) on how to use the NOMAD API.

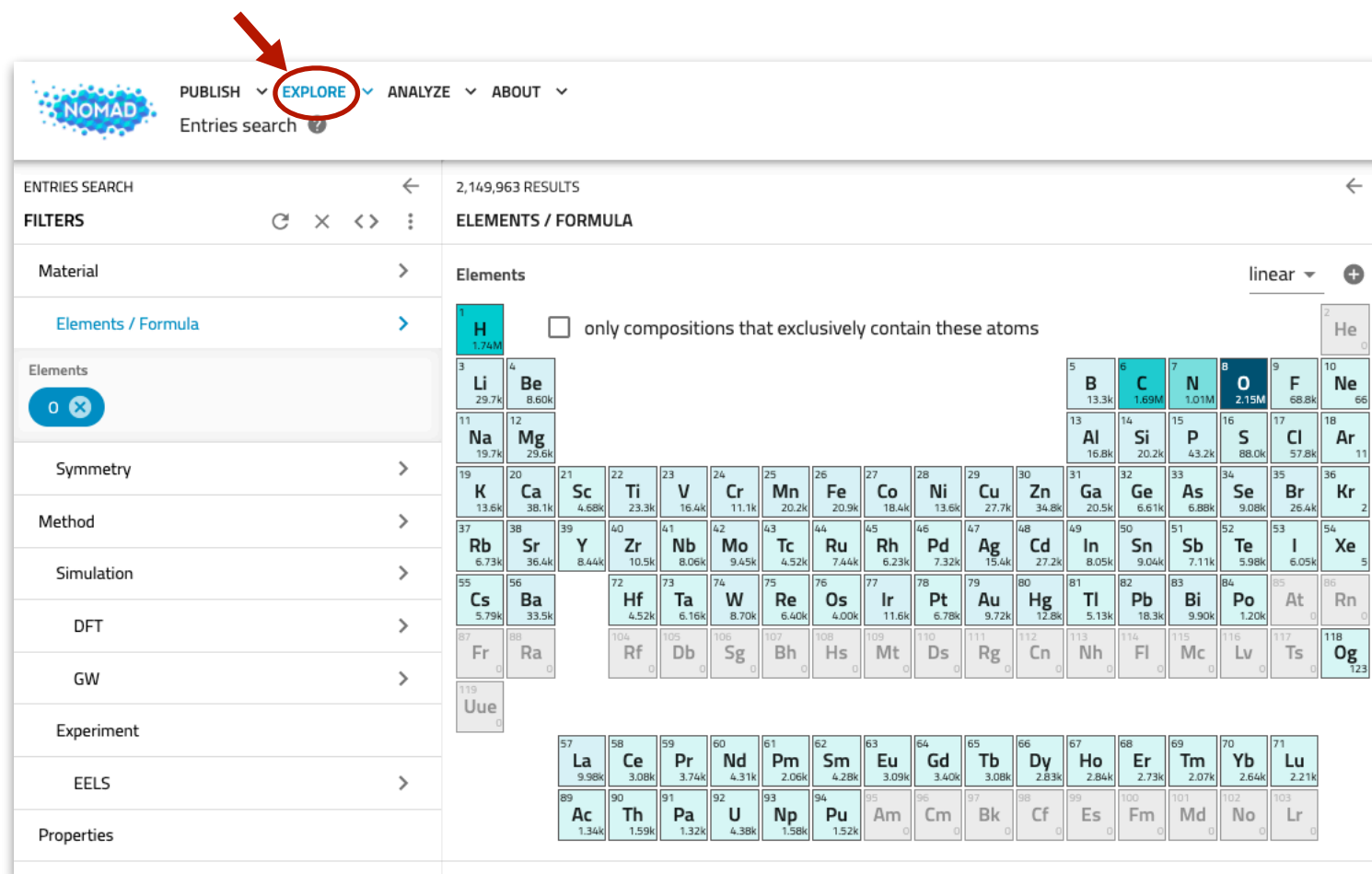
AI TOOLKIT TUTORIALS

There is a new version of NOMAD (1.0) that we currently provide as a beta version. This installation contains most of NOMAD's data and you can already use it to upload and publish more data. Eventually all data will be migrated to this version. It will become the official NOMAD after a short beta phase. We also provide an empty test version of NOMAD. You can use this to try the upload and publish process without any consequences. We will routinely void the test data.

**NOMAD 1.0 BETA**

NOMAD 1.0 TEST

# Querying the NOMAD Archive



**NOMAD** PUBLISH EXPLORE ANALYZE ABOUT

Entries search

ENTRIES SEARCH 2,149,963 RESULTS

FILTERS ELEMENTS / FORMULA

Material Elements linear +

Elements / Formula

Elements  only compositions that exclusively contain these atoms

0 x

Symmetry

Method

Simulation

DFT

GW

Experiment

EELS

Properties





Periodic table showing elements with counts: H (1.74M), Li (29.7k), Be (8.60k), B (13.3k), C (1.69M), N (1.01M), O (2.15M), F (68.8k), Na (19.7k), Mg (29.6k), Al (16.8k), Si (20.2k), P (43.2k), S (88.0k), Cl (57.8k), K (13.6k), Ca (38.1k), Sc (4.68k), Ti (23.3k), V (16.4k), Cr (11.1k), Mn (20.2k), Fe (20.9k), Co (18.4k), Ni (13.6k), Cu (27.7k), Zn (34.8k), Ga (20.5k), Ge (6.61k), As (6.88k), Se (9.08k), Br (26.4k), Rb (6.73k), Sr (36.4k), Y (8.44k), Zr (10.5k), Nb (8.06k), Mo (9.45k), Tc (4.52k), Ru (7.44k), Rh (6.23k), Pd (7.32k), Ag (15.4k), Cd (27.2k), In (8.05k), Sn (9.04k), Sb (7.11k), Te (5.98k), I (6.05k), Xe (5.73k), Cs (5.79k), Ba (33.5k), Hf (4.52k), Ta (6.16k), W (8.70k), Re (6.40k), Os (4.00k), Ir (11.6k), Pt (6.78k), Au (9.72k), Hg (12.8k), Tl (5.13k), Pb (18.3k), Bi (9.90k), Po (1.20k), At (0), Rn (0), Fr (0), Ra (0), Rf (0), Db (0), Sg (0), Bh (0), Hs (0), Mt (0), Ds (0), Rg (0), Cn (0), Nh (0), Fl (0), Mc (0), Lv (0), Ts (0), Og (123), La (9.98k), Ce (3.08k), Pr (3.74k), Nd (4.31k), Pm (2.06k), Sm (4.28k), Eu (3.09k), Gd (3.40k), Tb (3.08k), Dy (2.83k), Ho (2.84k), Er (2.73k), Tm (2.07k), Yb (2.64k), Lu (2.21k), Ac (1.34k), Th (1.59k), Pa (1.32k), U (4.38k), Np (1.58k), Pu (1.52k), Am (0), Cm (0), Bk (0), Cf (0), Es (0), Fm (0), Md (0), No (0), Lr (0).

# Querying the NOMAD Archive

PUBLISH ▾ EXPLORE ▾ ANALYZE ▾ ABOUT ▾

Entries search <sup>?</sup>

ENTRIES SEARCH 2,149,963 RESULTS

FILTERS    

Material >

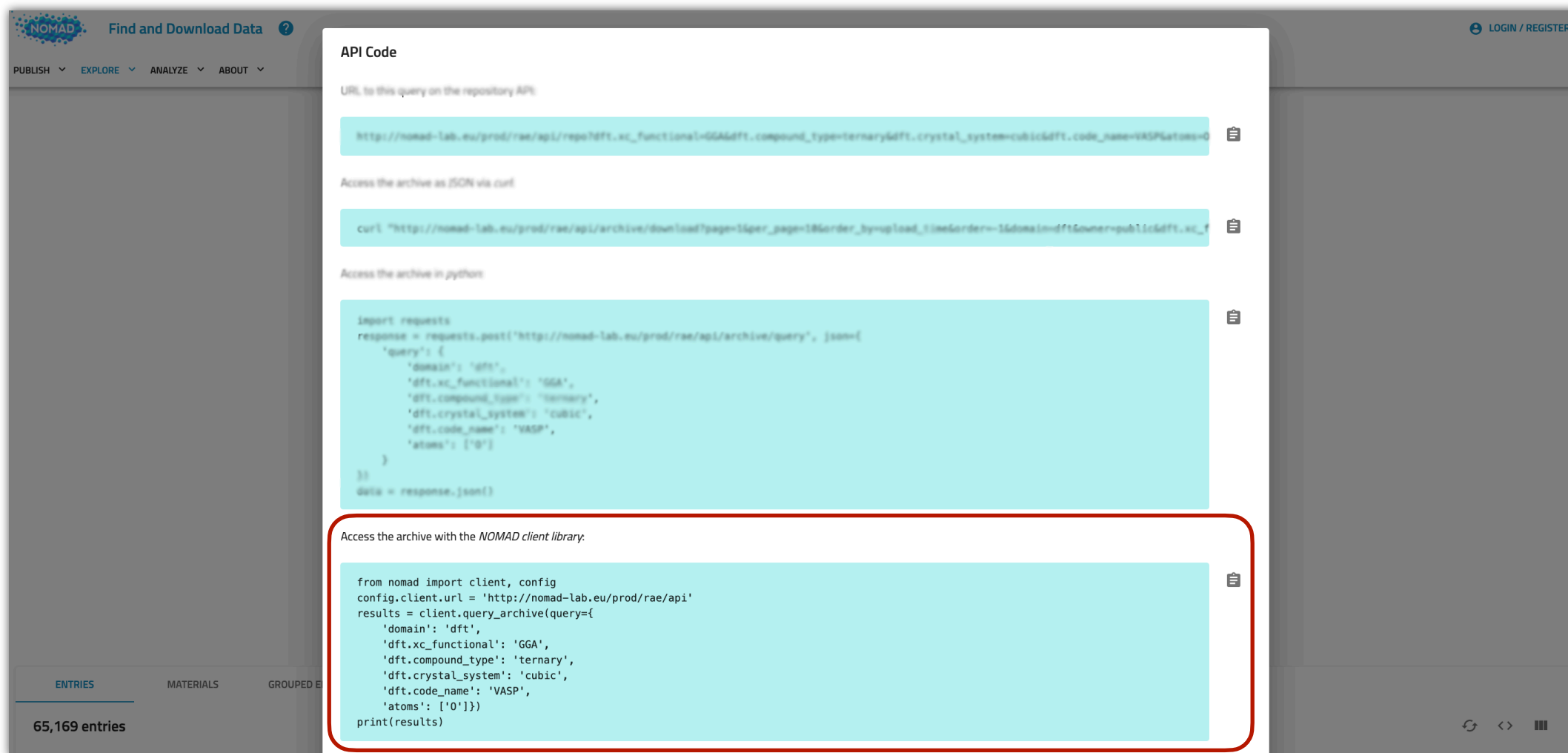
Elements / Formula >

Elements

only compositions that exclusively contain these atoms

1 H 1.74M																	2 He 0				
3 Li 29.7k	4 Be 8.60k															5 B 13.3k	6 C 1.69M	7 N 1.01M	8 O 2.15M	9 F 68.8k	10 Ne 66
11 Na 19.7k	12 Mg 29.6k															13 Al 16.8k	14 Si 20.2k	15 P 43.2k	16 S 88.0k	17 Cl 57.8k	18 Ar 11
19 K 13.6k	20 Ca 38.1k	21 Sc 4.68k	22 Ti 23.3k	23 V 16.4k	24 Cr 11.1k	25 Mn 20.2k	26 Fe 20.9k	27 Co 18.4k	28 Ni 13.6k	29 Cu 27.7k	30 Zn 34.8k	31 Ga 20.5k	32 Ge 6.61k	33 As 6.88k	34 Se 9.08k	35 Br 26.4k	36 Kr 2				
37 Rb 6.73k	38 Sr 36.4k	39 Y 8.44k	40 Zr 10.5k	41 Nb 8.06k	42 Mo 9.45k	43 Tc 4.52k	44 Ru 7.44k	45 Rh 6.23k	46 Pd 7.32k	47 Ag 15.4k	48 Cd 27.2k	49 In 8.05k	50 Sn 9.04k	51 Sb 7.11k	52 Te 5.98k	53 I 6.05k	54 Xe 5				
55 Cs 5.79k	56 Ba 33.5k	72 Hf 4.52k	73 Ta 6.16k	74 W 8.70k	75 Re 6.40k	76 Os 4.00k	77 Ir 11.6k	78 Pt 6.78k	79 Au 9.72k	80 Hg 12.8k	81 Tl 5.13k	82 Pb 18.3k	83 Bi 9.90k	84 Po 1.20k	85 At 0	86 Rn 0					
87 Fr 0	88 Ra 0	104 Rf 0	105 Db 0	106 Sg 0	107 Bh 0	108 Hs 0	109 Mt 0	110 Ds 0	111 Rg 0	112 Cn 0	113 Nh 0	114 Fl 0	115 Mc 0	116 Lv 0	117 Ts 0	118 Og 123					
119 Uue 0																					
57 La 9.98k	58 Ce 3.08k	59 Pr 3.74k	60 Nd 4.31k	61 Pm 2.06k	62 Sm 4.28k	63 Eu 3.09k	64 Gd 3.40k	65 Tb 3.08k	66 Dy 2.83k	67 Ho 2.84k	68 Er 2.73k	69 Tm 2.07k	70 Yb 2.64k	71 Lu 2.21k							
89 Ac 1.34k	90 Th 1.59k	91 Pa 1.32k	92 U 4.38k	93 Np 1.58k	94 Pu 1.52k	95 Am 0	96 Cm 0	97 Bk 0	98 Cf 0	99 Es 0	100 Fm 0	101 Md 0	102 No 0	103 Lr 0							

# Querying the NOMAD Archive



The screenshot shows the NOMAD website interface. At the top left, there's a navigation bar with 'PUBLISH', 'EXPLORE', 'ANALYZE', and 'ABOUT'. The main content area is titled 'API Code' and provides three methods to access the archive:

- URL to this query on the repository API:** A light blue box containing the URL: `http://nomad-lab.eu/prod/rae/api/repo?dft_xc_functional=GGA&dft_compound_type=ternary&dft_crystal_system=cubic&dft_code_name=VASP&atoms=0`
- Access the archive as JSON via curl:** A light blue box containing the curl command: `curl "http://nomad-lab.eu/prod/rae/api/archive/download?page=1&per_page=1&order_by=upload_time&order=-1&domain=dft&owner=public&dft_xc_f`
- Access the archive in python:** A light blue box containing a Python script using the `requests` library to query the API.
- Access the archive with the NOMAD client library:** A light blue box containing a Python script using the `nomad` client library, which is highlighted with a red border.

At the bottom left, there's a sidebar with 'ENTRIES' selected, showing '65,169 entries'. At the bottom right, there are navigation icons for refresh, back, and a menu.

# Querying the NOMAD Archive

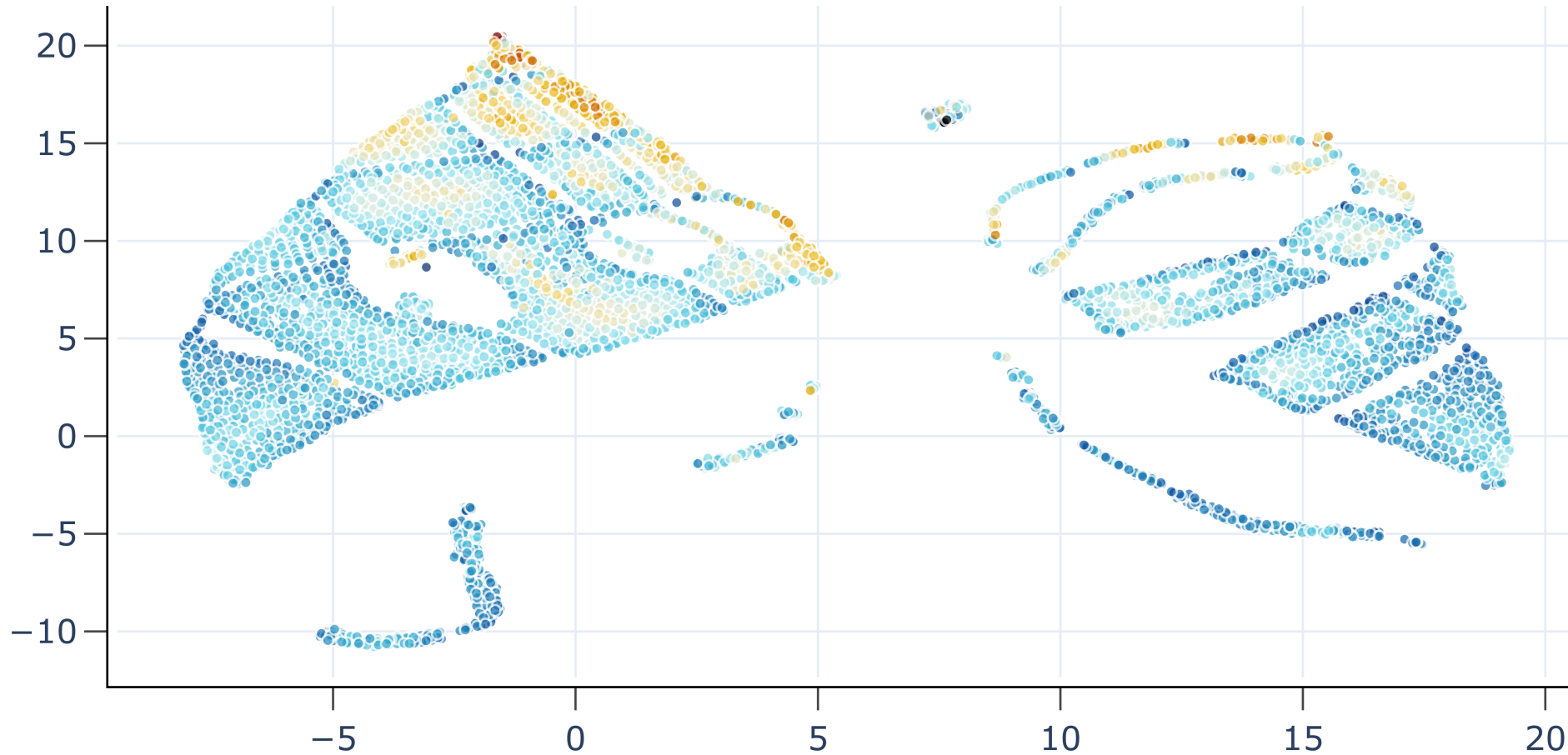


**Powerful Artificial-Intelligence Tools for Materials Science**

Find New Patterns and Information in Materials-Science Big Data

[Query the archive →](#) [View tutorials →](#) [Reproduce results →](#) [Get to work →](#) ⓘ

# Querying the NOMAD Archive



2-dimensional embedding, found with umap. Different colors means different atomic concentration values.



# Querying the NOMAD Archive



**Powerful Artificial-Intelligence Tools for Materials Science**

Find New Patterns and Information in Materials-Science Big Data

Query the archive →

**View tutorials →**


Reproduce results →

Get to work →


i



# AI tutorials for materials science



## Learn from tutorials



We develop and implement methods that identify correlations and structure in big data of materials. This will enable scientists and engineers to decide which materials are useful for specific applications or which new materials should be the focus of future studies. The following BEGINNER and INTERMEDIATE LEVEL tutorials are designed to get started with the AI Toolkit.

**Filter Tutorials**

Author ▾

AI Method ▾

System ▾

**BEGINNER LEVEL**

---

▾ **Symbolic regression via compressed sensing: a tutorial**

**Authors:** Emre Ahmetcik | Angelo Ziletti | Runhai Ouyang | Luigi Sbailò | Matthias Scheffler | Luca M. Ghiringhelli

---

▾ **Introduction to decision-trees methods**

**Authors:** Daniel Speckhard | Andreas Leitherer | Luca M. Ghiringhelli

---

▾ **Introduction to clustering**

**Authors:** Luigi Sbailò | Luca M. Ghiringhelli

---

- Tutorials span the most popular artificial-intelligence methods, such as clustering, dimension reduction, symbolic regression, decision trees, random forest, kernel ridge regression and deep neural networks.
- Applications to materials science are shown in a preparatory manner.
- Hands on exercises are presented throughout tutorials.
- Introductory videos, prepared in occasion of workshops and lectures held at the NOMAD Laboratory, are available.

# Querying the NOMAD Archive



**Powerful Artificial-Intelligence Tools for Materials Science**

Find New Patterns and Information in Materials-Science Big Data

-  [Query the archive →](#)
-  [View tutorials →](#)
-  [Reproduce results →](#)
-  [Get to work →](#) ⓘ

# Contribute

The *NOMAD Archive* can be queried from the *Artificial-Intelligence Toolkit*, and the data retrieved can be analysed on the fly with the desired AI tools.

The *Artificial-Intelligence Toolkit* hosts tutorials for hands-on learning of the most popular AI tools.

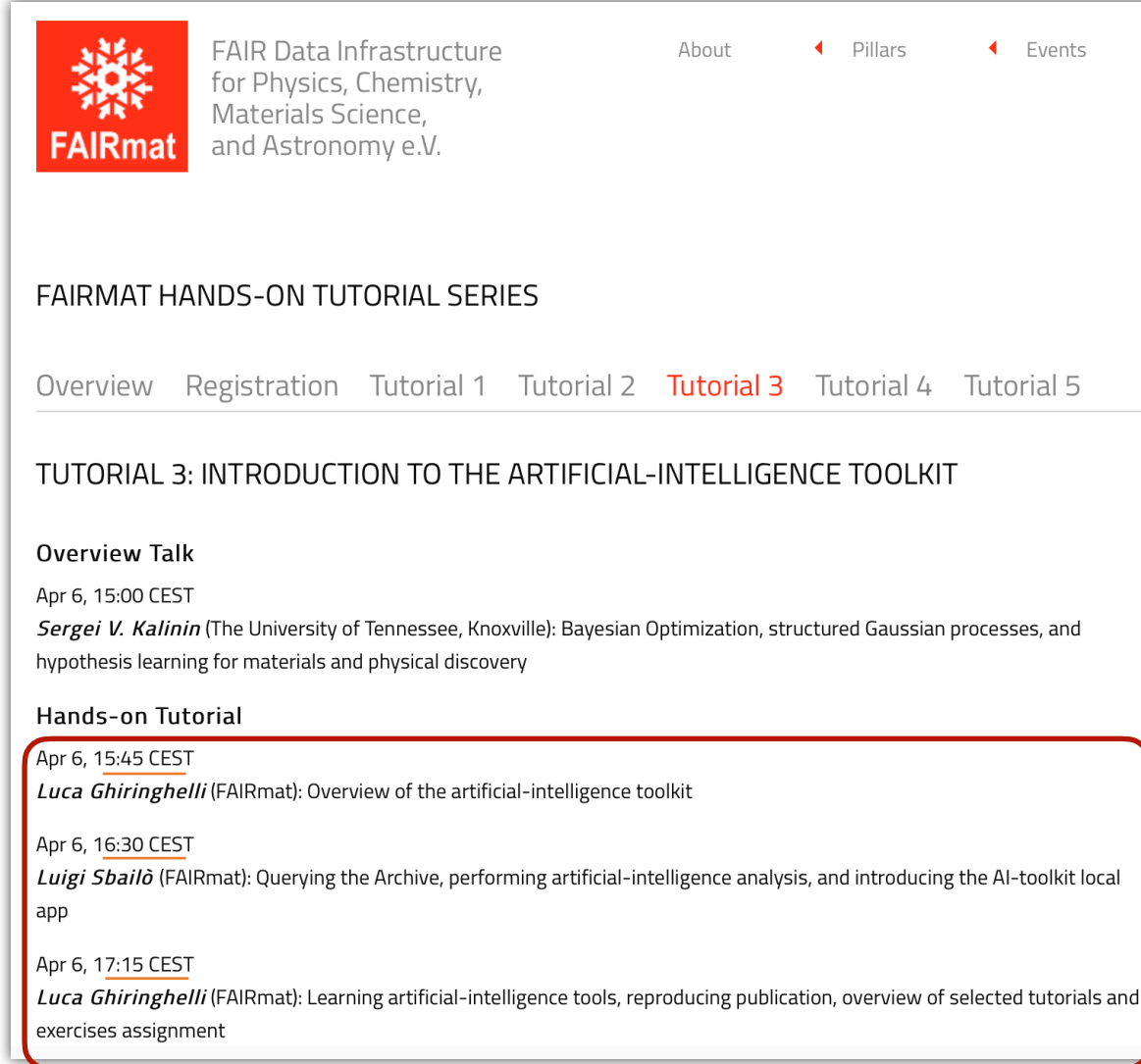
In the *NOMAD Artificial-Intelligence Toolkit*, we maintain several notebooks demonstrating the latest applications of AI to materials science data.

**Contact us!**

[ghiringhelli@fhi.mpg.de](mailto:ghiringhelli@fhi.mpg.de) [scheffler@fhi.mpg.de](mailto:scheffler@fhi.mpg.de)

# FAIRMAT HANDS-ON TUTORIAL SERIES

[www.fair-di.eu/fairmat-tutorials-3](http://www.fair-di.eu/fairmat-tutorials-3)



The screenshot shows the FAIRmat website. At the top left is the FAIRmat logo, a red square with a white snowflake and the text 'FAIRmat'. To its right is the text 'FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy e.V.'. In the top right corner are navigation links: 'About', 'Pillars', and 'Events'. Below the header is the title 'FAIRMAT HANDS-ON TUTORIAL SERIES'. Underneath is a horizontal menu with links: 'Overview', 'Registration', 'Tutorial 1', 'Tutorial 2', 'Tutorial 3' (highlighted in red), 'Tutorial 4', and 'Tutorial 5'. The main content area is titled 'TUTORIAL 3: INTRODUCTION TO THE ARTIFICIAL-INTELLIGENCE TOOLKIT'. It is divided into two sections: 'Overview Talk' and 'Hands-on Tutorial'. The 'Overview Talk' section lists a talk by Sergei V. Kalinin on April 6 at 15:00 CEST. The 'Hands-on Tutorial' section lists three activities on April 6: 1. At 15:45 CEST, Luca Ghiringhelli gives an overview of the artificial-intelligence toolkit. 2. At 16:30 CEST, Luigi Sbailò queries the Archive, performs artificial-intelligence analysis, and introduces the AI-toolkit local app. 3. At 17:15 CEST, Luca Ghiringhelli teaches how to learn artificial-intelligence tools, reproduce a publication, and overview selected tutorials and exercises assignment.

# Acknowledgments



Luca Ghiringhelli



Matthias Scheffler

Claudia Draxl, Christian Carbogno, Markus Scheidgen, Markus Rampp, Gian-Marco Rignanesi, James Kermode, Gábor Csányi, Ádám Fekete, Thomas A. R. Purcell, Mohammad-Yasin Arif, Lauri Himanen, Alvin Noe Lavines, Aakash Ashok Naik, Emre Ahmetcik, Andreas Leitherer, Daniel Speckhard, Marcel F. Langer, Benjamin Regler, Fawzi Mohamed, Angelo Ziletti