Ref. Ares(2022)707403 - 31/01/2022

HORIZON2020

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



D2.8

# Realization of INTERSECT data section on Materials Cloud

Flaviano J. dos Santos, Francesco Aquilante, Marnik Bercx, Arrigo Calzolari, and Nicola Marzari

www.intersect-project.eu

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



# **Document information**

Project acronym:	INTERSECT
Project full title:	Interoperable Material-to-Device simulation box for
	disruptive electronics
Research Action Project type:	Accelerating the uptake of materials modelling software
	(IA)
EC Grant agreement no.:	814487
Project starting / end date:	1 <sup>st</sup> January 2019 (M1) / 30th April 2022 (M40)
Website:	www.intersect-project.eu
Final version:	31/01/2022
Deliverable No.:	D2.8
Responsible participant:	EPFL (participant number 2)
Contributing Consortium members:	CNR
Due date of deliverable:	31/01/2022
Actual submission date:	31/01/2022
Dissemination level:	PU - Public

## Authors:

Flaviano J. Dos Santos, Francesco Aquilante, Marnik Bercx, Arrigo Calzolari, and Nicola Marzari

To be cited as:F. J. Dos Santos, F. Aquilante, M. Bercx, A. Calzolari and N.<br/>Marzari (2022): Realization of INTERSECT data section on<br/>Materials Cloud. Deliverable D2.8 of the H2020 project<br/>INTERSECT (final version as of 31/01/2022). EC grant<br/>agreement no: 814487, EPFL, Lausanne, Switzerland

# Disclaimer:

This document's contents are not intended to replace consultation of any applicable legal sources or the necessary advice of a legal expert, where appropriate. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user, therefore, uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view.

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



# Versioning and Contribution History

Version	Date	Modified by	Modification reason
v.01	18/01/2022	Flaviano dos Santos	First version
v.02	21/01/2022	Francesco Aquilante	First revision
v.04	26/01/2022	Arrigo Calzolari	Final version





# Content

1 Executive Summary	5
2 INTERSECT database within Materials Cloud	5
2. 1 Materials Cloud Archive	5
2.2 Materials Cloud three-dimensional crystals database (MC3D)	6
2.3 Materials Cloud Discover Section	7
2.4 Materials Cloud Explore Section	10
3 INTERSECT Internal Repository	11
4 Conclusions	15
References	16
Acronyms <sup>1</sup>	16

<sup>&</sup>lt;sup>1</sup> Acronyms are marked in purple in the text and defined at the end of the document.

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



# **1 Executive Summary**

This deliverable is a result of the activities carried out within Task 2.5 - *Data hub: Upscaling* Ginestra<sup>®</sup>*database* - and is dedicated to realization of an internal IM2D datahub, which collects the physical properties and data of materials generated by DFT codes within the IM2D infrastructure that can be used by Ginestra<sup>®</sup> in the device modeling step. The possibility of IM2D to access external databases within the OPTIMADE repositories has been already discussed in Deliverable D2.6 "Pipelines to/from external repositories"<sup>2</sup>) at M31.

As a starting point, we considered the data generated by the partners of the consortium during the INTERSECT activity. In this document, we discuss our policy to make these data publicly available. This consists of submitting data associated with our publications to the Materials Cloud Archive. Furthermore, we describe the *Materials Cloud three-dimensional crystals database* (MC3D), which encompasses part of our efforts to contribute to the material science community as well as for the future users of the IM2D simulation box and Ginestra<sup>®</sup>. This new materials database contributes with novel materials and in the diversity of methods used to compute the materials properties.

It is worth noticing that for general IM2D users, external to the INTERSECT project, the possibility to populate the IM2D datahub will remain a choice, with no obligations to make their data public. This will meet the needs of industrial users to protect their research and IP.

# 2 INTERSECT database within Materials Cloud

## 2. 1 Materials Cloud Archive

One of the biggest problems in science is currently facing is the so-called Replication Crisis [1]. In particular, research in computational science appears to suffer from the onset of this issue. Thus, many efforts in recent years have been spent to tackle this matter. For example, tools such as AiiDA help scientists to keep the complete provenance of their results. It has also become common practice to publish a collection of data associated with a given publication. This data usually contains extra information that cannot go inside the main text of the publication, but are essential for the full replication of the results. With this in mind, we strongly encouraged all partners to submit an entry to a public database storage associated with each of their publications related to the INTERSECT project.

Our repository of choice is the **Materials Cloud Archive** (<u>www.materialscloud.org</u>), an open repository for research data that are relevant to computational materials science. Materials Cloud Archive is mainly developed and maintained by EPFL, one of the consortium partners. Submissions to the Materials Cloud Archive receive persistent DOIs and a guaranteed lifetime of at least 10 years from submission. Experimental works are currently accepted only from authors affiliated with Materials

<sup>&</sup>lt;sup>2</sup> D2.6. https://drive.google.com/file/d/1EoKa1JxkqfCyM0jsIL1cq\_6YHOqOqKEn/view?usp=sharing

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



Cloud partners (INTERSECT included), or when the work relates to published computational results, e.g., by confirming or disproving them.

To submit a record to the Materials Cloud Archive, a user must register in the platform. This is needed so that one is able to (1) create new versions of records that have been accepted and published on the Materials Cloud Archive, (2) update their references and keywords, and (3) have access to your personal work area with all records you submitted to the Archive. After submission, the record will be reviewed by a moderator who may ask for clarifications or revisions, in order to uphold the standards of the archive, or even decline records deemed out of scope.

All partners have been asked to add a keyword "INTERSECT" to their records in the Materials Cloud Archive to make it easy to search and access all data related to the project in the Materials Cloud platform.

# 2.2 Materials Cloud three-dimensional crystals database (MC3D)

INTERSECT partners contributed to the realization of the MC3D database [2] that can serve the IM2D community but also the scientific community at large. Despite other materials databases having similar goals, such as the Materials Project, the MC3D database features novel materials and it is a push towards the diversification of materials databases. It is the first of its kind whose main quantum engine is Quantum ESPRESSO, one of the DFT codes included in the IM2D box. MC3D is hosted within the Materials Cloud database, and accessible through a public URL<sup>3</sup>.

The MC3D database [2] is a curated set of relaxed three-dimensional crystal structures based on raw CIF data taken from the external experimental databases MPDS [3], COD [4] and ICSD [5]. The raw CIF data received various layers of treatment: they were imported, cleaned and error corrected [6], filtered, and parsed [7,8] into a crystal structure. Only stoichiometric structures were kept into consideration.



duplications, as illustrated in Figure 1.

Figure 1. Venn diagram of the uniqueness analysis of all crystal structures (without partial occupancies) imported from the COD, ICSD and MPDS databases, after they were cleaned and parsed. The number represents the quantity of unique structures in each overlapping region [9].

Then, a uniqueness analysis was performed to avoid

<sup>&</sup>lt;sup>3</sup> <u>www.materialscloud.org/discover/mc3d</u>

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



The next step was to compute the ground state of all these compounds using density functional theory. For that, we used the SIRIUS-enabled [10] pw.x code of the Quantum ESPRESSO [11] distribution. We performed these calculations using tight tolerance criteria from the SSSP protocols [12]. This entire procedure is managed by AiiDA [13] workflows which automates the process while keeping full data provenance. Here, since the original source data of the ICSD and MPDS databases are copyrighted, only the provenance of the final SCF calculation on the relaxed structures can be made publicly available.

Currently, the database has about 35,000 compounds and it will be continuously updated with novel materials or improved calculations. We call "parents" the structures that we obtain from the initial screening of the several experimental databases, which search for new unique stoichiometric compounds. Once a parent structure has been optimized using density-functional theory, it is made public and added to the online Discover section of the Materials Cloud (copyrights might prevent publishing the original parent). The MC3D ID numbers come from a list of unique parents. Note that since not all structures have been calculated, some ID numbers are missing from the public version of the database.

The full ID of each structure also contains as an appended modifier the functional that was used in the calculations. Since the ID number points to the same unique parent, mc3d-1234/pbe and mc3d-1234/pbesol have the same starting point, but have been then relaxed according to their respective functionals.

# 2.3 Materials Cloud Discover Section

The Materials Cloud *Discover* Section presents a collection of curated research data. It features tailored visualization tools for the various data sets contributed by authors affiliated with the Materials Cloud partners, which includes the INTERSECT consortium.

The MC3D database can be accessed by clicking on its panel, as shown in Figure 2:



#### Materials Cloud three-dimensional crystals database (MC3D)

Authors: Sebastiaan Huber, Marnik Bercx, Nicolas Hörmann, Martin Uhrin, Nicola Marzari, Giovanni Pizzi

**Description:** Curated set of relaxed three-dimensional crystal structures based on raw CIF data from the experimental databases MPDS, COD, and ICSD.

## Figure 2. MC3D panel entry in the Materials Cloud Discover Section.

Users access the main MC3D page that includes an interactive search tool with a periodic table (Figure 3). Chemical elements that are present in any compound are colored in blue, elements in gray are not present in the databases.







Figure 3. MC3D search page. The search can be done by typing the compound name or by clicking on the elements that compose it.

The user can either search a compound by name or by selecting various elements in the periodic table. In the example below, we selected three elements, Mn, Fe, and Co. Notice that, as we go selecting some elements, many others turn to gray color. That helps to avoid combinations which are not present in the databases. In the example below (Figure 4), it results that the only compounds satisfying the filtering condition are CoFeGeMn and CoFeMnSi.

Next, the user can click on one of the results and explore it. For example, we chose to click on CoFeGeMn. We are then taken to a page that looks like Figure 5. We can see a *visual* representation of the crystal structure on the left-hand-side panel. A double-click on it will make the structure visualization interactive, where we can expand the unit cell, zoom, rotate, and more. In the central part of the page, there is the *Info* section, which contains the materials ID, formula, Bravais lattice and symmetry group information. The *Source* section keeps trace of the database that was the origin of the parent structure. The *Properties* section collects simulated physical properties, such as the total energy and magnetization of the material. Finally, on the right-hand side, information on the crystal structure cell and the atomic positions are given at 3D structure cell and 3D structure atomic coordinates sections.



Deliverable D2.8 Realization of INTERSECT data section on Material Cloud

← → C @ O A https://www.	materialsclo	ud.org/discov	er/mc3d/dash	iboard/ptable	2							ជ		© III\	ED 🕞 🛓	i 🔋 😵	•	<mark>©</mark> ∿ ≡
Ackn	iowledgem	ents												*				
E List																		
Search by name	н																	He
CoFeGemn	Li	Be											В	С	N	0	F	Ne
	Na	Mg											AI	Si	Ρ	s	CI	Ar
	к	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
	Rb	Sr	Y	Zr	Nb	Мо	Тс	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
	Cs	Ва	La	Hf	Та	W	Re	Os	lr	Pt	Au	Hg	ТІ	Pb	Bi	Po	At	Rn
	Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Nh	Fi	Mc	Lv	Ts	Og
				La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
				Ac	Th	Ра	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

Figure 4. Example of filtering search by selecting elements in the periodic table.

	Compound: CoFeGeMn			G	et data from RE	
Availabl	e space groups for this formula: F-43m (mc3d-46690/pbe)	v				
Structure 🔒	Info	3D structure c	ell			
Drag to rotate, scroll to zoom, right-click for other	MC3D-ID <sup>(info)</sup> : mc3d-46690/pbe Formula: CoFeGeMn		×[Å]	y[Å]	z[Â]	
Download structure:	Spacegroup international: F-43m	v1 0.	000000	2.852727	2.852727	
Y	Spacegroup number: 216 Bravais lattice: cF	v <sub>2</sub> 2	852727	0.000000	2.852727	
		V <sub>3</sub> 2	852727	2.852727	0.000000	
	Source	3D structure a	3D structure atomic coordinates			
		Kind label	x[Å]	y[Å]	z[Å]	
	Properties	Mn	2.852727	0.000000	-0.000000	
	Total energy: -14322.27 eV/cell 🝰	Fe	1.426364	1.426364	4.279091	
	Total magnetization: 4.02 µB/cell &	Co	1.426364	1.426364	1.426364	
	Absolute magnetization. 4.41 purceir 05	Ge	0.000000	0.000000	0.000000	
Dercell: 2 0 2 0 UPDATE RESET 2x2x2 CELL						
mera: x y z						

Figure 5. Individual compound explore page. An interactive visualization tool is offered for the crystal structure. Also, much information on the material is displayed.

It is worth mentioning that this database can also be accessed through the AiiDA REST API. This is a web-based protocol that allows us to automate the access to the MC3D database.

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



# 2.4 Materials Cloud Explore Section

Besides searching for the properties of individual materials and compounds through the *Discover* section of the Materials Cloud, users have the opportunity to browse the entire provenance of the database. This is provided by the *Explore* section of the Materials Cloud, which is an interactive browser for AiiDA provenance graphs (see D4.2 First Data management plan<sup>4</sup> and D4.6 Data management plan assessment and revision<sup>5</sup>).

AiiDA generates *data nodes*, which represent small units of information (such as crystal structures, input parameters, k-point meshes, codes, etc.) and *processes*, which are operations linking inputs, codes, and outputs. The *Explore* section presents three tabs (Figure 6). The first is the Grid tab, where one can list and browse through individual nodes and processes in the provenance graph. As illustrated below.

	O A https://www.materialscloud.org/explore/mc3d/grid/calculat				♥ IN @ ●	🗳 🔍 🗖 🍕	▶ 😋 🌆 😑
	DISCOVER EXPLORE ARCHIVE						More -
Explore Calculations							
Selected Profile: Material:	s Cloud three-dimensional crystals database	(MC3D)					
Data	Unique ID 🗧	Name	Creation time +	Last Modification time +	Creator	Process state	
Array     Bool	a95f38cc-ee40-484a-8f1a-ca33303edf18	PwCalculation	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
Code	acca6d2e-8e76-4b86-8893-ccaede56f31d	create_kpoints_from_distance	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
Dict	0ca77e2b-5602-467a-a193-8029bd505590	PwBaseWorkChain	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
Float	89814211-4c11-42c4-8e11-890ca46bf2ed	PwCalculation	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
FolderData	e1aee54c-dce2-41a6-863e-66a5cf582324	create_kpoints_from_distance	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
<ul> <li>Pseudo</li> </ul>	28b10b93-73c5-4f70-be49-eca25410379f	PwBaseWorkChain	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
► Remote	1ca1fdc5-2bd5-44d6-ac93-66d61d593535	PwCalculation	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
StructureData	775bd16b-6282-4708-8e6c-fcebf8f4fdc6	create_kpoints_from_distance	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
<ul> <li>Process</li> </ul>	fef2eb9c-9525-488f-adcc-ccf4653e494e	PwBaseWorkChain	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
Calculation     Workflow	8c39c924-b9df-40a3-b204-72c74c381076	PwCalculation	2 months ago	2 months ago	Sebastiaan Huber	FINISHED [0]	Details
Computer	* 1 2 3 4 5 6 7 10617 *					10 25 50	100 ALL

Figure 6. Initial page of the Materials Cloud Explore section for the MC3D database.

By selecting one of the entries in the database, users can click on *Details*, which automatically takes to the *Detail tab* parsing the uuid of the selected node. Here, we can visualize and download all the relevant files associated with the node, along with an interactive provenance representation of that node and the link to the provenance graph. See an example below (Figure 7), where a process is represented by a square, regular nodes by circles, and codes by triangles. Clicking on any of these nodes links to the corresponding *Detail* page. This allows one to navigate the entire provenance of a given node, calculation, or output.

<sup>&</sup>lt;sup>4</sup> D4.2. https://drive.google.com/open?id=1ZOqu25qmk2vTpDivEG6idHcwm0KYYly7

<sup>&</sup>lt;sup>5</sup> D4.6. https://drive.google.com/file/d/1WrwDRymnz6f1LPxtoQAmSlb98mmEC326/view

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



Finally, the Statistics tab gives access to a graphical summary of the data and activities related to the database, for example, the number of nodes of different types or the information on the calculations and the workflow execution.



Figure 7. The Details tab of the Materials Cloud Explore section.

# **3 INTERSECT Internal Repository**

A significant part of the work undertaken by the partners of the INTERSECT project has been dedicated to the study of materials with established and/or potential applications in synaptic electronics, whose constituent materials (e.g., amorphous chalcogenides, ferroelectric metal-oxides) are different from standard Si-based devices. The physical parameters of these systems are usually not available on materials databases. The goal is to collect and preserve the most significant findings of our research and make them available to the scientific community at large. Hence, our final aim is to make the data generated within INTERSECT publicly available.

Once any of the INTERSECT partners selects data for submission as part of a publication in a scientific journal, it has been internally agreed that they will also upload such data to the Materials Cloud Archive according to its specific policies, and with the inclusion of the keyword "INTERSECT" for ease of search. Figure 8 (top panel) shows an example of the "INTERSECT" query results on the Materials Cloud Archive and the corresponding published papers. Clicking on the single entries one has access to the publication details and the uploaded data (bottom panel).

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



aRepubblica LaBarc	caccia Biblioproxy TavolaPeriodica	Scholar WebPlotDigitizer Grain	ioundary Unit Conversions Dizionario	bab.la YouTube Aflowlib INTER	SECT T-one Plotly MP3Converte	er Omnia PDF_Drive
	🔂 Hom	0		🕁 Mi	aterials Cloud Archive	
LEARN	WORK DISCOVER	EXPLORE ARCHIVE				More
INTERSECT	Q About   Submiss	ion instructions   FAQs			Upload a record	og in 🛛 🐼 Sign up
ke	eywords	Found 5 results.	•	< 1 >	Sort by: Best mato	:h 😑
	intersact (4) max (4) max (4) smat (4) aida (2) aida (2) automated (2) common (2) dft (2) ere (2)	Hierarchical short- applications Coll 1934/constructors Taventi Francesco, Dia In the upcoming process materials for nonvolatile features of both the fugure status and the nonvolatile features of both the fugure results include the form results constructions the rom results construction the rom results constructi	and medium-range order st at Behnood. Catellari Alessandra, to ovecome the limitations of the si at Behnood. Catellari Alessandra, to ovecome the limitations of the si second the second second second second second second second second second second tattes at 1500 K and the anopolo- tist at a si at 1500 K and the anopolo- tist at a si at 1500 K and the anopolo- tist at the site of the second second second second tattes at 1500 K and the anopolo- tist at the site of the second second second second the Ge constraints and the second second second second standing of the interplay between ch standing of the interplay between the second second second second second second second second second seco	Catabani Arrigo andard von Neumann architectu Ini field, Ge-based compounds in Fisher (Sentandon Compounds in phase at 300 k of Ge-ich nai phase at 300 k of Ge-ich nai y na model of initiatomic poto closing synchronic struc- closing synchronic struc- tions synchronic struc- tions the mit mut Co-classical mic emilical composition, atomic stru- amical composition, atomic stru- opperties using different q der Jones, Dagermer Augustin.	e. x Se_1-x for selectors re, synaptic electronics is gaining have been proposed as switchin far dynamics, we study the struc- tar dynamics, we study the struc- tic polarization and restruct throug is polarizability of Sei ons. Our je olcal orders and investment throug becaute dynamics is multiations are a of Gielse chalcogenides. These cluve, and electrical properties in uuantum engines Dikan Vladimir, Einre Kristjan,	s ag nural nur.X h b 0 0 0 0 0
		Hignanese Gian-Marco,	Sewell Christopher J., Smit Berend,	, Tseplyaev Vasily, Uhrin Martin	Petretto Guido, Poncé Samuel, , Wortmann Daniel,	
● □ ~ < Repubblics LaBarca LEARN WOR	: > iccia Bibliogramy TavalaPeriodica BIK DISCOVER EXPLORE A	Rignanese Gian-Marco,	Sewell Christopher J., Smit Berend, archive materialiscloud org/record/202 dary Unit Conversions Dicionario Bab.	Tseplyaev Vasily, Uhrin Martin 22.7 Č) Ia YouTube Aflewilb INTERSECT	Petretto Guido, Poncé Samuel, , Wortmann Daniel, T-one Ploty MP3Converter Om	① ① +     nia PDF_Drive Books     Mo
E	ci > scria Bahrprey Tavalaferiodica isk DISCOVER EXPLORE AL Q. About   Submission Inst	Higmanese Glan-Marco, C units of the second	Sewell Christopher J., Smit Berend, archive.materialscloud.org/record/200 dary Unit Cenversions Dictionaria bab.	Tapiyaev Vasily, Ulwin Martin 22.7 C	Peretro Guido, Poncé Samuel, Wortmann Daniel, T-one Pisty MP3Converter Om Upticad a record	③ 宀 + nia PDF_Drive Books Mo
C      C     Republics     Laters     V     C	Construction of the second sec	Higmanese Glain-Marco, Construction of the second	Several Christopher J., Smit Berend archive materialecloud orghecord/2020 dary Unit Conversions Distancia bab	s Ge_x Se_1-x for	Pereterio Guido, Poncie Samuel, Wortmann Daniel, T-one Piety MP3Converter Om Uplead a record Expert	O     O     O     O
Compared	And Subservery TevelaPhrisdes     Biospray TevelaPhrisdes     And Subservery TevelaPhrisdes     And Sub	Higmanese Glaim-Marco, Description (Gamma) Schelar WestProttigtzer Orain Box Increve Increve Ange Catoose <sup>27</sup> Ange Catoose <sup>27</sup> Mars therait 10, Adorea, app corde, Scherood descellulationes, appear	Several Christopher J., Smit Bernd archive materialscloud orginecord/2000 day Unit Conversions Distance bab tructures in amorphous das califacilitation and prophoses	Tappiyaev Vasily, Uhrin Martin 22.7 C is YouTube Alfumb artEdisco s Ge_x Se_1-x for ont	Peretrio Guido, Ponce Samuel, Wortmann Daniel, T-one Ploty MP3Convertier On Uptred a recort Uptred a recort Control Data Control Data	Cord of the second of the
Conserved     Conserved	Comparison of the second comparison of th	Higmanese Glain-Marco, D Table Clain-Marco, Echair Westhictbirtzer Grain flow hochron PACS 7 Pacification PACS 7 Pacification PACS 7 Pacification PACS 7 Pacification PACS Pacification Pacifi	Several Christopher J., Smit Bernd a archive materialscloud orgineorid 2011 day Unit Conversions Distances betw thructures in amorphoue dra catalentifieren orzi, enge adalentifieren erden nange oder shockes in encogenous Gro	Tappiyaev Vasily, Uhrin Martin 22.7 C Is YouTube Attivetis INTERSECT S Ge_x Se_1-x for ov:1	Peretel outdo, Ponce Samuel, Vortrann Daniel T-one Piety MP3Converter Om Uptind a vector Uptind a vector Date Or Date Or	O     O
Consequence     Consequen	More and a second	Highneet Glain-Marco, Marcov Bodium - VestPercEpterer Grain Bounce Incover Incov	Several Christopher J., Smit Bernd archive.materialscloud orginecord/200 day Unit Conversions Distances Inth tructures in amorphocus descentrations of the several several descentrations of the several several results of the several several several results of the several	Taspayaev Vasily, Uhrin Martin      22.7      C      a YouTube Allowin INTERSECT      s Ge_x Se_1-x for      ord      s fa_i-s for advector applications, Mater      a reference dependence of applications,	Peretel outor, Pancel Samuel, Viormann Daniel,  Tone Patty MP3Converter Om  Uptrata recort  Uptrata recort  Export Date Or  Date Or  Inter Inte	Corporation     Corporati
Clarker     C	Comparison of the second	Highneed Gain-Marco, Highneed Gain-Marco, Scholar WebPutDigter Orale Box Accive Ac	Several Christopher J., Smith Bernd, archive, materialschool orgine certificit, day Unit Conversions Disnesses bab tructures in amorphouse etc. and an antibacture and an antipacture and etc. and antibacture and an antipacture and etc. and antibacture and an antipacture and etc. and antibacture and antipacture and etc. and antibacture and antipacture and etc. and antipacture and antipacture and antipacture antipacture and antipacture and antipacture antipacture and antipacture antipacture and antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture antipacture an	Tappiyoev Vaasiy, Uhrin Martin      22.7      (c)      a YouTube Altweb antESECT      s Ge_x Se_1-x for      ord      x Be_1-s for advector applications, Mare      a development of a-memory comparing the      down and a second applications, Mare      a development of a-memory comparing the      down and a second applications, Mare      a development of a-memory comparing the      down and applications, and applications, Mare      a development of a-memory comparing the      down and applications, and applications, Mare      a development of a-memory comparing the      down and applications, application	Norman Daniel Tone Paty MP3Converter On Uglass awart Cone Paty MP3Converter On Uglass awart Date Or Date Or Date Or Norman	O D + +
Clarker     C	Market State	Highneek Gain-Marco, Highneek Gain-Marco, Schulz WebPutDigtzer Orale Box Accivity	Several Christopher J., Smit Bernd, archive, materialschool orgineerer?2000 day. Unit Conversions: Disnesses lab tructures in amorphous: the calibration of the calibration of the calibration of a calibration of the structures in an explosed readow encodes a spanning excitation of the matter encodes a spanning excitation of the structures of the structures in an explosed and the matterials and the structures in an explosed and the readow encodes and the structures in the calibration and the structures are structures and the structures	Taspiyaev Vaasiy, Uhrin Martin      22.7      (c)      a YouTube Altweb antESECT      s Ge_x Se_1-x for      ord      x Se_1-s for advector applications, Mare      a development of in-memory school/applies      bring of the formula bit has a school in the memory school/applies	Norman Daniel Tone Ploty MP3Converter On Uptical avent Uptical avent Export Date Ce	O D + +
Contraction     Contracti	Control of the second sec	Highnanese Glain-Marco, Ministry VeshPetrolipticer Grain Bouncer Incourse	Several Christopher J., Smit Bernd archive.materialschuud orgineeroof 2000 day Unit Conversions Distances Inth tructures in annorphocus des catalaceitance or it, engs catalaceitance weber and product structures in encodence method structures and annorphocus excellent of the structures in encodence method structures in annorphocus excellent of the structures in encodence in disconductions, in generating in the total in disconductions, in generating in the total in disconductions, there each is the structures in disconductions. These structures in the outpeople Description	Tappiyoov Vaasiy, Uhrin Martin      22.7      C      a YouTube Afford NTERSECT      s Ge_x Se_1-x for      orat      cont     a development of a spectra spectra spectra spectra      orat      a development of a spectra spectra spectra spectra      orat	Neretto Guido, Poncie Samuel,	Cordent Former Sealers      Cordent Former      Cordent Former Sealers     Cordent Former Sealers     Cordent

Figure 8. Top) Screenshot of the "INTERSECT" query in the Materials Cloud Archive. Bottom) Screenshot of selected entry details and data.

However, since part of the INTERSECT results has not been published yet (scientific papers are in preparation), the unpublished data cannot be uploaded on the materials cloud archive, i.e., made public. For this reason, each partner agreed to gather the data, generated within the project but not yet published, on a private INTRANET storage session, accessible only to the consortium members, through the INTERSECT website<sup>6</sup>. Such an internal database is conceived as a dynamic entity that

<sup>&</sup>lt;sup>6</sup> <u>https://intersect-project.eu</u>

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



partners populate along with the production of data relevant for the upcoming publications. Once published, data will be added to the Materials Cloud Archive, as INTERSECT outcomes (Figure 8). The general structure of the database is simple and not meant to be optimal for search and retrieval of data, but allows one to have a view of the ensemble of data produced or still in production within INTERSECT.

Each partner has complete freedom in selecting the amount and type of data that they deem relevant for a future publication. Most likely, the intranet database will include a larger amount of data than the one finally submitted to the Materials Cloud Archive, as the process of publication often requires a selection of data across a long time-span and cannot be predicted with enough certitude. In this respect, the intranet database is also an internal tool to keep track of the overall data produced during the project. Each member of the INTERSECT project can access data produced and stored in the partner's entry of the intranet database. The frontpage to the database indicates the partner ownership, as seen in the Figure 9:

		://intranet.intersect-project.eu/index.php/apps/files/?dir=/INTERSECT%20Database&fileid=2522		
	Files			
	All files	INTERSECT Database		
*	Favorites	Name 🔺		
<	Shared with you	AMAT	arrigo.calzolari	000
<	Shared with others	CNR	< arrigo.calzolari	
S' Q	Shared by link	EPFL	< arrigo.calzolari	000
		FMC	< arrigo.calzolari	
		FRA	arrigo.calzolari	000
		ICN2	arrigo.calzolari	
		МЕС	arrigo.calzolari	
		7 folders		

Figure 9. Snapshot of the internal INTERSECT database, stored in the INTRANET section of the INTERSECT web portal.

Although the data within each partner's folder is organized as to the owner's convenience, the typical structure includes subfolders where, for example, data are associated with a specific material or physical process. As an example, we show the entry from CNR group (Figure 10), where data are organized in terms of physical systems, e.g., crystalline GeSe, amorphous GeSe, metal/chalcogenide interfaces (top panel). Specific subsystems are organized in subdirectories, for example per dopants, dopant level, and stoichiometry (bottom panel).





=	Files	WIERSECT Intranet	
	All files	INTERSECT Database      CNR      T	
*	Favorites	Name 🔺	Size
<	Shared with you	a-GeSe <a>arrigo.calzolari</a>	•••• 18.7 MB
\$	Shared with others	C-GeSe C-GeSe	••• 111.5 MB
S O	Shared by link Tags	C Interfaces	•••• 646 KB
	-	3 folders	130.9 MB
=	Files	<b>NTERSECT</b> Intranet	
	All files	$\begin{tabular}{c c c c c c } \hline \begin{tabular}{c c c } \hline \begin{tabular}{c c c } \hline \begin{tabular}{c c } \hline $	
*	Favorites	Name 🔺	Size
<	Shared with you	dos.ps <a>arrigo.calzolari</a>	•••• 67 KB
<	Shared with others	Ge.dat	•••• 33 KB
ø Q	Tags	Ge40Se60_Si-10%.cif	•••• 41 KB
	C C	nbnd-Occupation-Energy-IPR.dat	•••• 71 KB
		plot-band.gnu	•••• < 1 KB
		Se.dat	••• 33 KB
		Si.dat <a>arrigo.calzolari</a>	•••• 33 KB
		TDOS.dat	•••• 54 KB
		8 files	333 KB

Figure 10. Example of folder structure organization of the internal database, per each project group (CNR in this case).

The data available for this particular system have been selected by the authors on the basis of general criteria, mostly in order to provide full reproducibility of the results (e.g., crystal structure, software-specific parameters for setting up the calculation, etc.), as well as the access to processed data for analysis of the physical properties of the system (e.g., density of states, band structure, etc.).

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



# **4 Conclusions**

The present deliverable describes the effort by the INTERSECT partners to ensure that the data produced during the project will be stored in a reliable and easy-to-access way for the benefit of the scientific community at large. Each partner is committed to upload onto the Materials Cloud Archive the data relevant to all its INTERSECT-related publications. In the deliverable, the policies of such uploading are described that guarantee full reproducibility of the published results. Meanwhile, a database restricted to the INTERSECT partners has been set up that contains data related to research that has not yet been published. Besides its usefulness as backup storage, the role of such a database is that of giving an overview of the data produced within the project before the typical "skimming" that is required by the time of the actual publication. Each partner has access to all data from the project stored on this internal database, with no restriction for the format used, and very general guidelines for selecting the data. Finally, we also described the *Materials Cloud three-dimensional crystals database*, currently with about 35,000 materials known experimentally. The database also contains the ground-state properties of the materials obtained with high-throughput calculations automated with AiiDA. This is an effort for the diversification of materials databases, which is also very useful for the user of the IM2D simulation box developed by our consortium.

Deliverable D2.8 Realization of INTERSECT data section on Material Cloud



# References

[1] John P. A. (August 1, 2005). "Why Most Published Research Findings Are False". PLOS Medicine. 2
(8): e124. doi:10.1371/journal.pmed.0020124. ISSN 1549-1277. PMC 1182327. PMID 16060722.

[2] https://www.materialscloud.org/discover/mc3d/dashboard/ptable#aboutWindow

[3] The Pauling File <u>http://paulingfile.com/</u> exposed through the Materials Platform for Data Science <u>https://mpds.io/</u>.

[4] S. Gražulis et al. Crystallography open database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. Nucleic Acids Research, 40:D420–D427, 2012, <a href="http://www.crystallography.net">http://www.crystallography.net</a>.

[5] Inorganic Crystal Structure Database, http://www.fiz-karlsruhe.com/icsd.html.

[6] A. Merkys et al. COD::CIF::Parser: an error-correcting CIF parser for the Perl language Journal of Applied Crystallography 49 (2016)

[7] S. P. Ong et al. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. Computational Materials Science, 68:314 – 319, 2013.

[8] A. Togo. Spglib. http://spglib.sourceforge.net.

[9] S. Huber, et al. In preparation.

[10] SIRIUS, https://github.com/electronic-structure/SIRIUS.

[11] P. Giannozzi et al. Advanced capabilities for materials modelling with Quantum ESPRESSO. Journal of Physics: Condensed Matter, 29:465901, 2017.

[12] SSSP protocol for the calculation of structural and thermodynamical properties of inorganic materials, Nicolas Hoermann et al., to be published.

[13] G. Pizzi et al. AiiDA: Automated Interactive Infrastructure and Database for Computational Science. Computational Materials Science, 111:218-230, 2016. <u>http://www.aiida.net</u>.

# Acronyms

**API** - Application Programming Interface

DFT – Density Functional Theory

IM2D – Interoperable Materials to Device

MC3D - Materials Cloud three-dimensional crystals database

**OPTIMADE** - Open Databases Integration for Materials Design

- **REST** Representational State Transfer
- UUID Universally Unique Identifier