

# Case Study 3: Combining data sources in Materials Physics

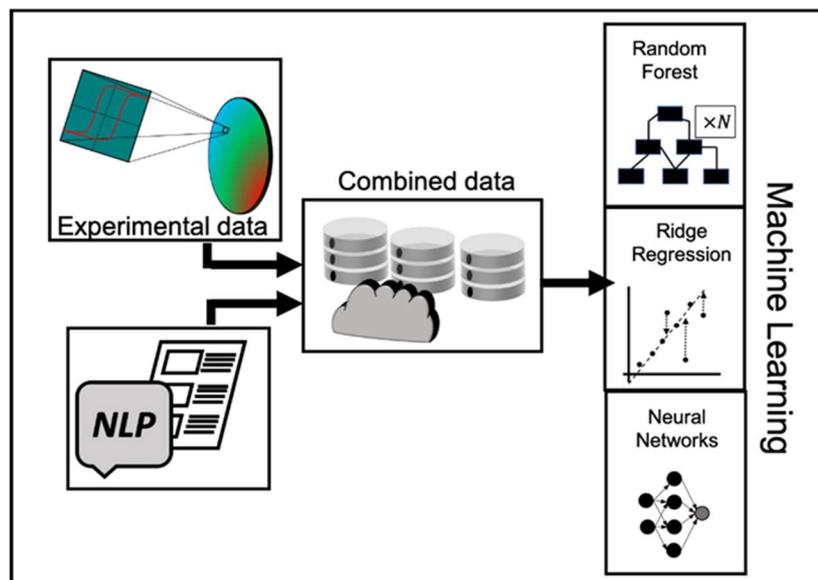
*Dr Richard Rowan-Robinson<sup>1</sup>, Dr Zhaoyuan Leong<sup>1</sup>, Prof Nicola Morley<sup>1</sup>, Dr Noel Vizcaino<sup>2</sup>, Vasily Bunakov<sup>2</sup>*

<sup>1</sup>*Department of Materials Science and Engineering, University of Sheffield, Sheffield, S1 3JD*

<sup>2</sup>*Science and Technology Facilities Council, Scientific Computing Department. Rutherford-Appleton Laboratory. Oxfordshire. OX11 0QX*

## 1 Introduction

Inspired by the success of the field of Bioinformatics, Material Informatics aims to utilise the vast data resources available to discover new functional materials. However, the data so central to this approach, although vast, is largely unstructured and uncoordinated.



*Figure 1. Schematic of the case study.*

Unstructured data is buried within journal article text, figures and tables and needs to be extracted to form tabulated structured databases of material properties, which are machine readable and from which insights can be drawn. Experimental data from industrial and academic partnerships is uncoordinated, where even with the best intentions, a lack of standardised conventions means the raw unprocessed data is often squandered after publication, stored locally on hard disks, and forgotten once the primary researchers move on.

CS3 aims to understand the challenges around merging data sources within the field of Material Science, both the unstructured data within the literature and the uncoordinated data continually generated within the laboratory (Figure 1). The work focuses on the Material Science subdomain of functional magnetic materials, for which there is ongoing research in the

Morley lab. The first component explores existing tools for harvesting data from the literature. Here, the focus is on the open source ChemDataExtractor software, since this has already been demonstrated to be successful for extracting the Curie temperature of magnetic materials and the MatScibert software, which is specifically designed for Material Science. Due to issues arising with compiling MatSciBert, we also wrote a Mathematica code aiming to extract the composition and magnetic properties from the abstract corpus we put together. The potential for generalising these tools for extracting other magnetic material properties will be explored and recommendations made in light of the challenges in applying this methodology. It is expected that the use of Natural Language Processing (NLP) tools such as ChemDataExtractor will be indispensable for realising the untapped potential of the data embedded within the literature.

Alongside this, CS3 will establish the requirements of a combined data architecture for storing NLP harvested data alongside the experimental data measured within the Morley laboratory. Here, the foundations of an RDF (Resource Description Framework) database will be developed. These graph databases store data as a network of linked objects which will enable complex queries to be utilised to extract grouped/connected data from the database. The study explores the potential of automatic serialisation to standardise the data output format (e.g. serialisation to JSON-LD format for storing in the RDF database).

## 2 Main Outputs

The outcomes are split across two components: first, the application of NLP to extract data from the function magnetic material literature subdomains, and secondly, the merging of this data with laboratory data as an RDF database. Each part is discussed in the report, with the example data sets etc given in the appendix.

### 2.1 Data Extraction Using NLP

The project initially determined the different open source NLP software available (Table 1).

Table 1: Current list of NLP software available for materials science data extraction.

Software name	Programming Language	Link to software
ChemDataExtractor	Python	<a href="http://chemdataextractor.org/">http://chemdataextractor.org/</a>
pyNLTK	Python	<a href="https://www.nltk.org/">https://www.nltk.org/</a>
OpenNLP	Java	<a href="https://opennlp.apache.org/">https://opennlp.apache.org/</a>
OSCAR4	Java	<a href="https://github.com/BlueObelisk/oscar4">https://github.com/BlueObelisk/oscar4</a>
MatScibert	Python	<a href="https://arxiv.org/abs/2109.15290">https://arxiv.org/abs/2109.15290</a>
Stanza (Python version of Stanford CoreNLP)	Python	<a href="https://stanfordnlp.github.io/stanza/">https://stanfordnlp.github.io/stanza/</a>

From this table, two different NLP softwares were investigated to determine their suitability for extracting magnetic properties from literature. The two softwares were: ChemDataExtractor and MatSciBert. ChemDataExtractor was chosen as it had already been used to extract the Curie Temperature of magnetic materials [<https://www.nature.com/articles/sdata2018111>], and was a python language code that we could adapt for magnetic properties. MatSciBert was chosen because it utilises the state-of-the-art BERT NLP model, with some specialisation pretrained for the Material Science domain. It is also useful to compare ChemDataExtractor, which utilises primarily manually tuned rule-based NLP against MatSciBert which utilises neural networks. Further a Mathematica code was written using Regular Expression in combination with heuristics method, due to issues that arose in the compiling of MatSciBert. In the Appendix are screenshots of the Jupyter notebook (Figure A1) and the Mathematica code (Figure A2) used during this part of the project.

The first action was to develop a small corpus of eight texts on a range of different magnetic materials (Table A1 in the Appendix provides a list of the papers used). From these papers, the aim was to extract the following magnetic properties: saturation magnetisation ( $M_s$ ), coercive field ( $H_c$ ) and Curie temperature ( $T_c$ ).

The data was manually extracted so that the accuracy results of the NLP software could be assessed. The two NLP softwares were then adapted, where in the case of ChemDataExtractor, new parsing rules were added to enable it to extract magnetic properties. For the MatSciBert code, we defined a set of labels that would be used by the software to be assigned to each word in the text, we then chose the labels which were linked to the magnetic parameters to be extracted from the text. Our Mathematica code used a hybrid NLP-Regular Expression methodology, which was written to extract the required parameters and compositions from the abstracts, with the aim to link the composition to the different magnetic parameters. The TextSentences function built into Mathematica was used to break up the text into individual sentences and Regular Expressions were used to identify terms of interest. A logical heuristics was coded in to identify parameters with compositions of interest.

Table 2 summarises the effectiveness ( $\epsilon$ ) of the two NLP softwares and our Mathematica code to extract these properties. The full outputs from the software are found in Table A2, A3 and A4 in the appendix. The effectiveness is given by:

$$\epsilon = \frac{\text{number of extract parameters}}{\text{total number of possible parameters to extract}} \times 100$$

Table 2. Summary of the results from the two different NLP software

Pa per	Chemdataextractor					MatSciBert					Mathematica code				
	Com	$M_s$	$H_c$	$T_c$	$\epsilon$ %	Com	$M_s$	$H_c$	$T_c$	$\epsilon$ %	Com	$M_s$	$H_c$	$T_c$	$\epsilon$ %
[1]	✓	✓	✓	-	100	x	x	x	-	0	✓	✓	✓	-	100
[2]	✓	✓	✓	✓	80*	x	x	x	x	0	✓	✓	✓	x	75
[3]	x	✓	-	-	50	x	x	-	-	0	✓	✓	-	-	100
[4]	In	✓	-	-	75	x	x	-	-	0	✓	✓	-	-	100

	part														
[5]	x	✓	✓	-	66	x	x		-	0	✓	✓	✓	-	75**
[6]	In part	✓	-	-	75	x	x	-	-	0	✓	✓	-	-	75**
[7]	✓	✓	-	-	100	x	x	-	-	0	✓	✓	-	-	75**
[8]	✓	✓	-	-	100	x	x	-	-	0	✓	✓	-	-	100

\* an additional magnetisation was extracted that didn't exist in the abstract; \*\*the composition and the saturation magnetisation were both extracted, but weren't linked together

From Table 2, it is observed that the Chemdataextractor is excellent at outputting the magnetic properties for abstracts, but struggles to assign a composition to these properties, if they are not written in a certain way within the text. While our Mathematica code could extract both the composition and magnetic properties, but only ~60% of the time link the two together. This would be the next step in the development of the software to allow for over 90% effectiveness in retrieving the composition and linking it to the magnetic properties. Due to the limited amount of time we had on the project, the work on MatSciBERT was incomplete. It is observed in Table A4, that the outputs from the programme were rubbish, and further iterations of training the software are required to achieve sensible outputs.

The Chemdataextractor codes used within the Jupyter notebook are found on the github site linked to this case study.

## 2.2 Merging functional magnetic material datasources

For this part of the case study, we produced a datamap of measurements following a typical magnetic sample (Figure 2). A typical data set for each of the different experiments listed in Figure 2 are provided in the Appendix.

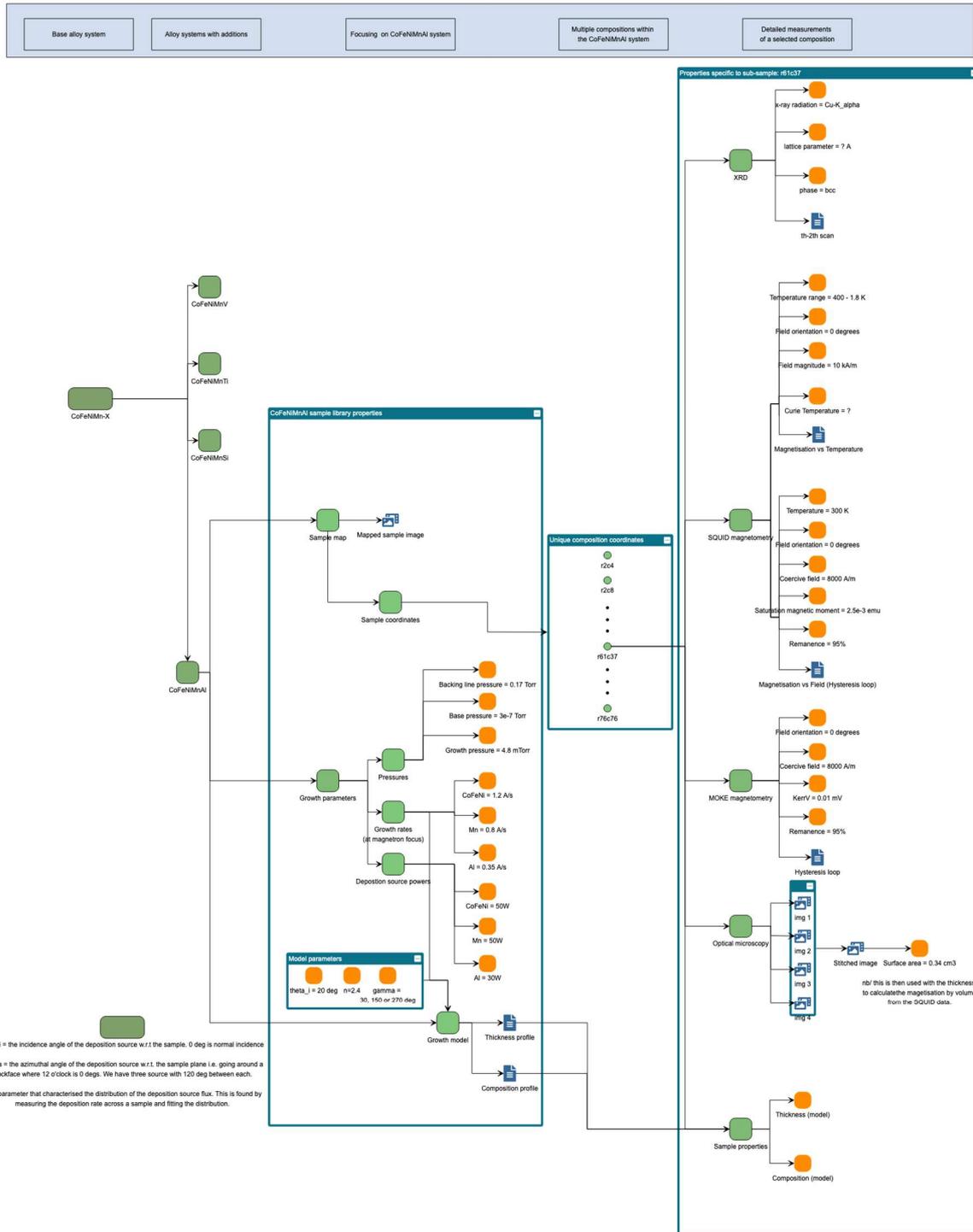


Figure 2. UoS experiment tree for merging different data outputs

The data is for the film CoFeNi<sub>50</sub>W<sub>Mn</sub><sub>30</sub>W<sub>Al</sub><sub>30</sub>W, which was fabricated using combinatorial sputtering on a 3-inch wafer (Schematic of the experiment given in Figure A3). This means that across the 3-inch wafer there are over 180 different compositions. The theoretical compositions across the wafer are calculated using in-house software written in Python (output film: Figure A3b). The wafer is first photographed (Figure. A4), then mapped into 2 mm squares. This coordinate system is then used through the rest of the characterisation for the structural and functional properties, and link them back to the predicted composition at that

point. Thus the data merging tools for the experimental data have to be able to read the different data files, and link the different properties to each of the co-ordinates. Two of the characterisation measurements are high-throughput experiments (MOKE, XRD), which mean that the wafer is cut into 1-inch strips and placed within the equipment. The equipment then uses the coordinate system (Figure. A4) to take data at set intervals across the wafer, such as the magnetisation hysteresis loop (Figure. A5) and the XRD spectra (Figure. A6). While other characterization techniques (SEM, SQUID) require the wafer to be cut into smaller 4 mm squares and the data taken associated with the coordinate point. Further the MOKE software is programmed to automatically read the coercive field and max voltage from the hysteresis loops to produce a ternary plot for the 180 compositions (Figure A7). When merging all this data into one database, to gain the most out from it, the software must be able to read different forms of data, and link that data back to the corresponding coordinate on the wafer.

An example dataset from one sub-sample within the larger sample library with coordinates row 61, column 37 was serialised to a JSON-LD format ready for integration into a RDF database. Each data set at the present is manually added to the RDF database, by the user. Also the majority of the computational and experimental work was recorded in a Jupyter notebook. This electronic notebook is owned by the user, and unless shared with others, it is private and can not be accessed.

## 3 Issues that arose during the case study

### 3.1 Data Extraction Using NLP

For the data extraction using NLP, a number of different issues arose due to the software not being specifically designed for magnetic materials, and therefore having to be adapted to perform the data extraction. Also the NLP software had limitations even after being changed, in part due to the algorithms used and in part due to the magnetic properties and compositions being extracted not being in the same form in every text. Plus we had issues compiling the MatSciBERT software, so it was not very user friendly, and required a lot of training plus a supercomputer to run.

The main issues that arose were:

- Many magnetic properties are a function of other parameters e.g. magnetization is a function of temperature, maximum entropy change is a function of applied magnetic field, power loss is a function of field frequency. This requires more complex data structures beyond property-value pairing since these properties depend on an additional variable.
- Remanent magnetization is sometimes quoted as a percentage or a magnetization. To convert between them requires knowing the saturation magnetization. Varying terms, such as residual magnetization are also used.
- Magnetic properties are often measured as part of investigations of superconducting materials. Here  $T_c$  generally refers to the superconducting transition temperature, but can be misidentified as the Curie temperature in the context of magnetic materials.
- The alloy compositions are written in different formats, and the actual composition was not always extracted, sometimes only part of it was returned.
- Many properties can be extracted from figures e.g. coercive field, saturation magnetisation and remanent magnetisation can all be obtained from a hysteresis loop.

However, unless these parameters are explicitly quoted in the text or a table there are no existing automated extraction tools to obtain this information from figures.

### **3.2 Merging functional magnetic material datasources**

For the merging of the different data sources into the RDF database a few issues arose during the process. These included:

- The NLP outputs can be programmed to be automatically inputted into the RDF database, but each input has to be manually checked to ensure that it is correct and there are no errors, this takes time.
- For the experimental data, the inputting of the data files strongly relies on the user, as it is not automatically done by the different experimental techniques. To remove the emphasis on the user, having all the systems automatically connected to the database, would ensure a smoother method to merge the data. Also reduces the human error within the process.
- This can be a challenge for older experimental systems, for which the control software is reliant on an old operating system to function. This is a reasonably common occurrence and means many control PCs are not connected to the internet due to their operating system not being supported any longer, which could present difficulties for a direct pipeline of data from the experiment to the RDF database.

## **4 Benefits and recommendations for the future**

The conclusions from this case study are that the proof of concept is good, meaning that NLP can be used to data mine literature for properties, but at the present there are too many limitations to be useful for magnetic materials. Thus the NLP algorithms have to be either further adapted to cover the issues that arose during the case study (section 3) or new specific NLP algorithms be created, which are designed specifically for magnetic materials, and are designed to resolve the issues that arose in the adapted NLP software.

Although some infrastructure is available, it is in its infancy and the documentation associated with it is poor. This means that it is not easy for a novice in the area to adapt/change the algorithms to suit their research field for data extraction. For this form of materials informatics to expand into different materials research fields, the documentation associated with the software needs to be improved, to allow less experienced users to adapt it for their research areas. For example, these types of project are becoming more popular with undergraduate students who have basic programming knowledge, so the documentation should be written at their level.

Using computational software such as NLP and machine learning (ML) provides cheaper, quicker and simpler methodologies for database creation and materials discovery. It is likely that this range of tools would benefit a broad range of material science applications. Training the next generation of researchers in both computational and experimental skills sets will be important, as academia and industry look to benefit from the Materials Informatics “tool set”. One way to achieve this would be to put together a CDT, where the first year would be used to train the PhD students within the different computational algorithms and experimental techniques, then years 2-4 of the PhD would involve them using these skills to solve user case research problems, being based within different research groups and working with industry.

## 5 Required infrastructure

From the outcomes of this case study the main infrastructure requirements for being able to combine different data sources (NLP data sets, experimental data sets) to achieve one structured database containing all data are:

- User friendly tools for non-developers.
- Better documentation for existing tools.
- Tools for extracting data from figures.
- Common database system where data is automatically uploaded to and assigned the correct metadata.
- Joint sharing of electronic notebooks for collaboration, to allow groups to develop code etc together.

## 6 Data Strategy

### 6.1 Introduction

CS3 takes part of a global **data strategy** for data holdings, as delineated briefly by CS4. This means that all data, metadata and arising dataflows share the same foundational standards. It is FAIR metadata by design. FAIR data will be strongly promoted if not enforced. Steps can be taken to make data FAIR, beyond original data.

The metadata blueprints are based on semantic linked data technologies. Core standards are **W3C DCAT 2** and **W3C JSON-LD 1.1**. There will be many other standards involved.

The metadata was developed in tandem with CS4 efforts. CS3 metadata is compatible by design, as well with other DCAT profile.

#### 6.1.1 CS4 Compatibility

These sections from CS4 related to metadata are equally applicable to CS3:

- Whole **Section 3 Data Strategy** (except that the scientific additions for CS3 will be different, including dataset types).
- Entire **Appendix A**.
- Entire **Appendix C** (Albeit the *specifics for CS3* that we will mention are not included there but here).
- CS3 and CS4 metadata serialisations are compatible and complementary by design. (A coherent and methodical approach must be followed).
- CS3 and CS4 metadata serialisations will have a degree of compatibility with DCAT2 (and thus **Dublin Core**).

We will focus mostly on CS3 particularities and perhaps covering any gaps left by CS4 (where applicable). All is a brief overview, hinting to future work.

#### 6.1.2 Data Strategy Summary

FAIR stands for Findable Accessible Interoperable and Reusable.

We can summarise the data strategy main points:

- Data Strategy includes FAIR goals.
- Extensive use of international standards.

- Separation of metadata and data.
- Use of Permanent Identifiers (PIDs) whenever possible.
- Data is supplied externally or internally.
- Most of the metadata is generated automatically from the data.
- Rich metadata. Further enrichment can be incremental.
- Highly structured metadata (by design) and data (from raw data, if possible).
- Reuse of existing metadata structures.
- Minimisation of user input at any stage.
- Mandatory **UTF-8** encoding (without BOM) for any text supplied.
- The original tool is the best source of its data (and its conversions).
- Supplied data must use using open standards if possible.
- Self-documented entries (with domain considerations).
- In the absence of native metadata repository support, a metadata manifest file (named for all instances e.g., **PSDI-MANIFEST.jsonld**) can be included along with the data for ingestion in any persistence system.
- **JSON-LD** serialisations are both an **RDF** and **JSON** document. Therefore, three processing options are available. (Cited by expressivity order).
- Besides standards recommendations. Following best practices e.g., “*things not strings*” philosophy:
  - JSON-LD. <https://w3c.github.io/json-ld-bp/>
  - Semantic data. <https://www.w3.org/TR/cooluris/>
  - Data on the Web. <https://www.w3.org/TR/dwbp/>
- To avoid a tower of babel of ontologies. Before creating new “things” it is best to reuse definitions from existing widespread standards.

### 6.1.3 Systems context

To illustrate the various aspects, we will use the graph from the experiment for the film **CoFeNi\_50W\_Mn\_30W\_Al\_30W** to illustrate emerging dataflows and their I/O.

(Note: data package supplied with that *folder name*).

There will be options to describe the experiment and other entities with precision in a machine processable-friendly way. Not all metadata will be added at once. It is more likely to be followed by a long tail of minute additions.

The main idea is to take the data processing and knowledge representation away and let researchers concentrate on their tasks.

## 6.2 The Knowledge Graph (KG)

A goal of the metadata is to produce and audit trail from the dataflows (which we understand as all possible paths for data in a network). Usually, dataflows follow a Directed Acyclic Graph (DAG) structure, but we are not going to completely rule out cycles just yet (feedback loops). We may abstract them away. We can reserve the workflow term for higher level procedural flows.

The idea is to encode the experiment, its stages and all data and metadata I/O. It is open to additions, and it should be descriptive enough. Derived data or sometimes mid-step temporary data should be offered from encoded data, if convenient.

The modelling of processes and their audit logs could be encoded using the **W3C PROV-O**, a provenance ontology standard. This means datasets will also be entities part of an activity (by humans or machines). This is also amenable to be called *workflow*, more concerned about the big picture (e.g. including the recording of instrument use, human task).

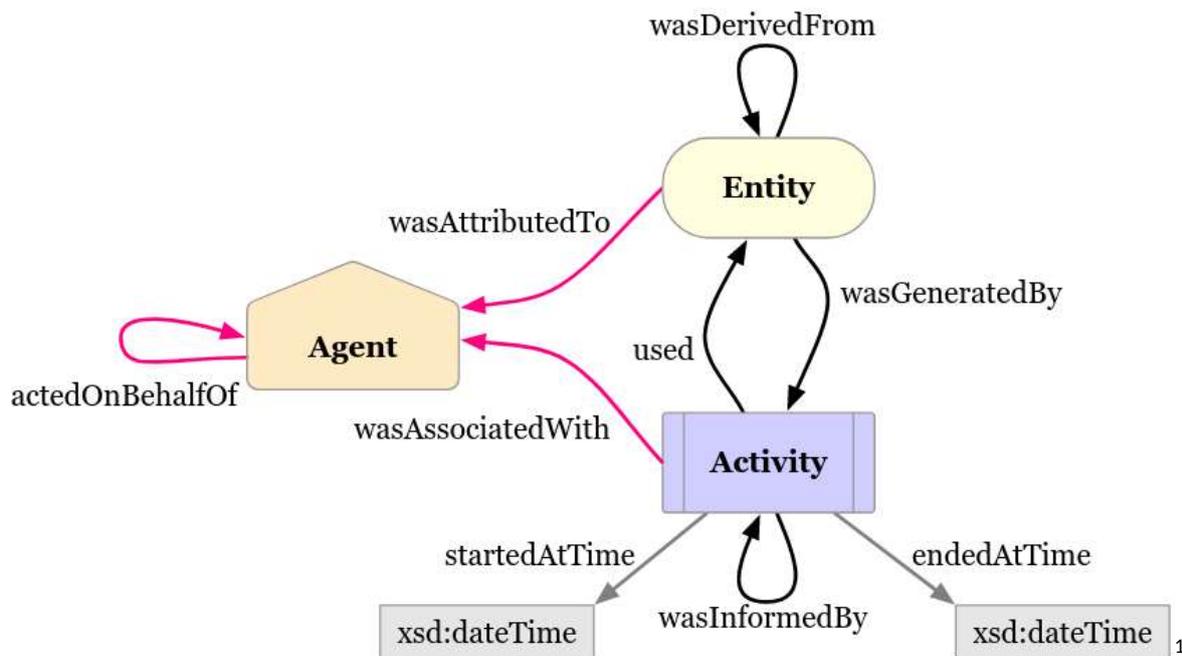


Figure 3. Basic sketch of PROV-O from <https://www.w3.org/TR/prov-o/>

We want to capture the knowledge of the experiment nature itself in a way that can automate most data processing and be shared (also, with the world).

We have an Experimentation (`prov:Activity`) with an Experiment dataset (a `dcat:Dataset` but also a `prov:Entity`). But DCAT has provisions to deal with this. We can research the best way to model things across PSDI.

The experimental command line tool uploaded to CS4 GitLab can create dataset metadata from the contents of folders or zip files. We can create a *Sample* dataset per sample (region as provided e.g., r61c37) and then aggregate them into a *OpticalMicroscopy and Wafer* dataset.

## IRIs

The nodes (vertices) and the links (edges) or the graph are identified, sometimes externally. Some of them will be IRIs which are internationalised URIs. IRIs can be designed to be PIDs.

<sup>1</sup> Basic sketch of PROV-O from <https://www.w3.org/TR/prov-o/>

## PIDs

They will follow the recommended PID scheme (guideline):

<internet-domain(base)><type><concept><reference>

The essential idea is they should be permanent with no added information that could change.

It should be noted that the JSON-LD semantic context will add information to the UUID to make a Permanent Identifier (PID) e.g. (type N/A):

- For datasets metadata (as *DCAT datasets*): id = <base>/dataset/<UUID>
- For data files metadata (as *distribution views*): id = <base>/distribution/<hash digest>

These ideally would give online access to the metadata they refer to (dereferenceable PIDs).

IRIs preferred in any case. Other examples are DOI, ORCID, ror.org, ISBN, ISSN, etc.

## 6.3 Datasets

The key JSON-LD node of this scientific Knowledge Graph is the dataset. CS3 has its own type of datasets, and their additional metadata could be quite different from CS4. Of course, this requires careful analysis, but we can give hints of realistic solutions. Datasets are part of a web of metadata, never in isolation and never final.

We can abstract things away so anything can be a dataset. If we need to bundle information or need any of the DCAT dataset properties, then we may have a **dataset candidate**. Any data files are a clear giveaway.

Some observations:

- Anything with files (input or output) likely to be a dataset
- Module I/O likely to include parameters at dataset level

Entities	Range	dct:type	Tweaks and observations
Experiment	dcat:Dataset, prov:Entity	Experiment	Initial Input parameters. Resulting output parameters (may include datasets).
Wafer	dcat:Dataset, prov:Entity	76mm diameter. Wafer. Subtypes according to their layer of data (view): <ul style="list-style-type: none"><li>• FilmComposition</li><li>• KerrVoltage</li><li>• OpticalMicroscopy</li></ul> Series. Any other series at dataset level.	Output. Collection of Samples. Different samples parameters per wafer type.
Sputtering (Stage)	dqv:hasQualit yMeasuremen t	N/A	Initial parameters

Growth Parameters	dqv:hasQualityMeasurement	N/A	Initial parameters
XRD	dcat:Dataset, prov:Entity	XRD	Output. Collection of Strips.
SQUID Magnetometry	dcat:Dataset, prov:Entity	SQUID	Output.
High-throughput MOKE	dcat:Dataset, prov:Entity	HTMOKE	Output. Collection of Strips. <ul style="list-style-type: none"> <li>magnetisation hysteresis loop (and PDF)</li> <li>Ternary data and PDF.</li> </ul>
Strips (or Slices)	dcat:Dataset, prov:Entity	Wafer, Strip	Output. Collection of Samples (from physical strips). 76mm/3 width.
Sample	dcat:Dataset, prov:Entity	Sample (to identify quantifiable differences). e.g, Optical 2mm, SQUID 4mm,	I/O (e.g., output within Optical Microscopy)

A dataset like a series of samples can be bundle into a wafer dataset series with the parameters specific for the sample. We may decide later at what level is appropriate add the parameters (same or wafer). The case for making it a dataset may be warranted by the existence of output files (distribution) or any other metadata property making it suitable to be a dataset.

### 6.3.1 Experiment Dataset

Following our machine processable schemas, the experiment:

- It will have an ID (great as folder/bucket name too)
- Free keywords (tags)
- Various formal categorisation options.
- A title (for users, not machines)
- A detailed description (for users, not machines)
- Authors and contributors, etc.
- Groups (e.g., Projects)
- Organisations
- Temporal information (when and for how long)
- Spatial (“where”, quite optional but it is easy so why not and remember the coriolis force, reproducibility, context)
- DOI, PURL
- Associated papers if any (also DOIs)
- Citations, etc.
- Context parameters (known) e.g.:

- CoFeNi ? Mn? Al? Proportions? etc
- Power applied to the layers
- Long etc.

A trusted source of information for the elements could be:

- [https://pubchem.ncbi.nlm.nih.gov/rest/pug/periodictable/JSON/?response\\_type=display](https://pubchem.ncbi.nlm.nih.gov/rest/pug/periodictable/JSON/?response_type=display) (Download available in JSON, XML format)
- <https://iupac.org/what-we-do/periodic-table-of-elements/> (IUPAC) **Not processable.**
- **Desired:** Any RDF/OWL representation, if available.

The formats were never used poorly (even if XML standard is quite old!) and there are massive knowledge representation gaps. The idea is to describe things linking to further things and let the system decide the **level of depth** required.

## 6.4 I/O Parameters

To encode I/O parameters **W3C Data Quality Vocabulary** is a prominent solution among institutional standards. We can express quality dimensions defined in **ISO/IEC 25012**.

For datasets is an entry in *hasQualityMeasurement*, but they can be on their own. We can define our custom metrics.

Example entry for sample region params:

```
"hasQualityMeasurement": [
  {
    "waferRow": {
      "@type": "QualityMeasurement",
      "value": 61,
      "isMeasurementOf": {
        "waferCoordinate": {
          "@type": "Metric",
          "inDimension": "precision",
          "expectedDataType": "xsd:unsignedShort"
        }
      }
    }
  },
  {
    "waferColumn": {
      "@type": "QualityMeasurement",
      "value": 37,
      "isMeasurementOf": {
        "waferCoordinate": {
          "@type": "Metric",
          "inDimension": "precision",
          "expectedDataType": "xsd:unsignedShort"
        }
      }
    }
  },
  {
    "coordinateLength": {
      "@type": "QualityMeasurement",
```

```

    "value": 2,
    "isMeasurementOf": {
      "millimetre": {
        "@type": "Metric",
        "inDimension": "precision",
        "expectedDataType": "xsd:unsignedShort"
      }
    }
  }
}
]

```

## SI Units

Instead of defining the SI units, it is best practice to use a (trusted) vocabulary. QUDT (vocabulary) was originally developed by NASA and later adopted by many prominent standards. It is aligned with ISO 80000 standards for units and quantities. It seems a suitable choice for our DCAT profile and used by other international standards.



e.g., instead of the custom metric “millimetre” this can be done via “*unitMeasure*”: “<http://qudt.org/vocab/unit/MilliM>” an **SDMX** standard property (**recommended**).

```

{
  "coordinateLength": {
    "@type": "QualityMeasurement",
    "value": 2,
    "isMeasurementOf": {
      "unitMeasure": "http://qudt.org/vocab/unit/MilliM"
    }
  }
}

```

## 6.5 Tools and processes

The tools information, whether is from a physical instrument or software, should be included as its own dataset or as a dataset addition. A tool is not a dataset but the information package regarding the tool is. There will be a **registry of tools** across PSDI. There could be classed as internal (in-house) or external. Online and offline.

### 6.5.1 Tool descriptive information

An example of an external tool produced data would be e.g., “**Diffraction system=XPRT-3**” as identified in their metadata XRDML files (<https://www.malvernpanalytical.com/en/products/product-range/xpert3-range>). There is plenty of metadata information to be mined from them (based on relevance).

Their XRDML files follow XML XSD schemas provided in (their `landingPage`):

<https://www.malvernpanalytical.com/en/products/category/software/x-ray-diffraction-software/data-collector/#specs>

Example data supplied by University of Sheffield uses the schema version 2.0.

The tool (or other tool) creates the CSV files that will feed a script (XRD-plots.vsz) to generate a PDF report. The PDF is a view suited for (only) human consumption of the data in the CSV files. These could be two distinct datasets representing information layer and presentation layer.

## 6.5.2 NPL tools

NPL Software tools use can be added as a manual or automated step (whichever applies, depending on systems integration level).

Let us illustrate the process for a NPL tool (category) from the tool's registry:

1. Inputs
  - a. The paper(s) to be analysed is recorded as a dataset.
  - b. Keep any raw data.
  - c. Add convenient raw data conversion(s), if relevant.
  - d. The processing parameters.
  - e. Bundle all as a **dataset input** (in practice, a dataset network).
  - f. Reuse the tool for similarly packaged inputs (pluggability).
2. Process
  - a. Attribute using the id from the NPL software (tools repository).
  - b. Add User(s) attribution.
  - c. Add Organisations.
  - d. It could include temporal and spatial params. Where, when, duration.
  - e. Add project and related metadata.
  - f. Add to an experiment (and/or experiment stage).
3. Outputs
  - a. Keep raw output
  - b. Add a machine friendly conversion(s).
  - c. Bundle all as a **dataset output** (in practice, a dataset network)
  - d. Reuse the tool for similarly packaged outputs (pluggability).

This modelling could slightly vary, and it is not comprehensive.

It should be noted datasets e.g., may have parent or child datasets or be associated with anything. Due to the **Linked Data** interconnected nature, datasets are always part of larger web of metadata.

## 6.5.3 Jupyter Notebooks

A Jupyter notebook is just a text file, including source code entries and past outputs. We could treat it as a module/process unit with a concise function. Furthermore, it can be broken down into its cells (as chained processing steps). They could be added to the tool repository (likely as internal tools). The process for all modules should be quite like the one described for NLP.

## 6.6 Data access

Based on this example (distribution view for a data file):

```
{
  "id": "9c46fe262c300499e863c2600b952570",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img2.jpg",
  "accessURL": "9c46fe262c300499e863c2600b952570",
  "downloadURL": "9c46fe262c300499e863c2600b952570",
  "mediaType": "image/jpeg",
  "byteSize": "126106",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "9c46fe262c300499e863c2600b952570"
  },
  "conformsTo": []
}
```

(Note: Not comprehensive, e.g., licensing info on a per file basis could be included)

A public download URL for a data file could be:

<https://www.psd-domain.com/download/9c46fe262c300499e863c2600b952570>

If the domain is <https://www.psd-domain.com/>

It should be noted that the JSON-LD processor will help **building the PID using the semantic context e.g.:**

- API access (internal or external) `accessURL` = <domain>/data/<hash-digest>
- For web clients (HTTP GET) `downloadURL` = < domain>/download/<hash-digest> or any web clients.

Bases with their domains may be included in the semantic context.

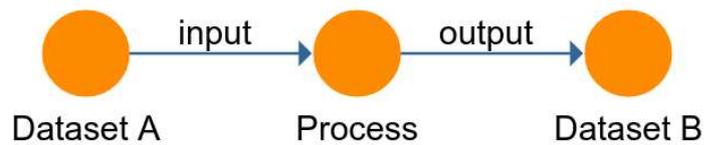
In the case a tool requires a data file from a dataset the file:

- It will be in the distribution property as an entry.
- The relative path within the package is included in filename (SPDX standard).
- It will be identified by a hex hash digest (e.g., MD5).
- It is likely to persist in object storage (if not meant to be queried, otherwise it will be ingested the RDF aware database). This includes small text `mediaType` files.

## 6.7 Special Attribution provisions

e.g., software or processes in a DCAT dataset

We may have dataset inputs (likely with files) or other inputs.



We must attribute the process in dataset B as the creator, e.g.:

```
"qualifiedAttribution": [  
  {  
    "type": "Attribution",  
    "agent": {  
      "type": "SoftwareAgent",  
      "id": "our-process-ID",  
      "name": "eg-prediction.py"  
    }  
  }  
]
```

More commonly, a person, group, or organisation with a **well-defined role** from a vocabulary (standard) e.g., ISO 19115-3, but any other appropriate one can be used:

```
"qualifiedAttribution": [  
  {  
    "type": "Attribution",  
    "hadRole": "funder",  
    "agent": {  
      "type": "Organization",  
      "id": "https://ror.org/0439y7842",  
      "name": "Engineering and Physical Sciences Research Council"  
    }  
  }  
]
```

## Appendix A

Table A1: List of abstracts used to test the NLP software

Paper	
[1]	Pavithra, C.L.P., Janardhana, R.K.S.K., Reddy, K.M. <i>et al.</i> An advancement in the synthesis of unique soft magnetic CoCuFeNiZn high entropy alloy thin films. <i>Sci Rep</i> 11, 8836 (2021). <a href="https://doi.org/10.1038/s41598-021-87786-8">https://doi.org/10.1038/s41598-021-87786-8</a>
[2]	Ma Y, Wang Q, Zhou X, Hao J, Gault B, Zhang Q, Dong C, Nieh TG. A Novel Soft-Magnetic B2-Based Multiprincipal-Element Alloy with a Uniform Distribution of Coherent Body-Centered-Cubic Nanoprecipitates. <i>Adv Mater.</i> (2021), 33(14), e2006723. doi: 10.1002/adma.202006723.
[3]	Cramer CL, Nandwana P, Yan J, Evans SF, Elliott AM, Chinnasamy C, Paranthaman MP. Binder jet additive manufacturing method to fabricate near net shape crack-free highly dense Fe-6.5 wt.% Si soft magnets. <i>Heliyon</i> , (2019), 5(11):e02804. doi: 10.1016/j.heliyon.2019.e02804..
[4]	Xu Y, Zhu Z, Zhao H, Zhou J. Preparation of $\epsilon$ -Fe(Si) <sub>3</sub> N Powder Using a Salt Bath Nitriding Reaction and a Study on the Soft Magnetic Properties of the Powder. <i>Materials</i> , (2019);12(2):228. doi: 10.3390/ma12020228.
[5]	Zhu M, Fa Y, Yao L, Tao P, Jian Z, Chang F. The Influence of Annealing on the Structural and Soft Magnetic Properties of (Fe <sub>0.4</sub> Co <sub>0.6</sub> ) <sub>79</sub> Nb <sub>3</sub> B <sub>18</sub> Nanocrystalline Alloys. <i>Materials</i> , (2018), 11(11):2171. doi: 10.3390/ma11112171.
[6]	Zhai S, Wang W, Xu J, Xu S, Zhang Z, Wang Y. Effect of Co and Gd Additions on Microstructures and Properties of FeSiBAlNi High Entropy Alloys. <i>Entropy</i> , (2018), 20(7):487. doi: 10.3390/e20070487.
[7]	Wang J, Li J, Wang J, Bu F, Kou H, Li C, Zhang P, Beaugnon E. Effect of Solidification on Microstructure and Properties of FeCoNi(AlSi) <sub>0.2</sub> High-Entropy Alloy Under Strong Static Magnetic Field. <i>Entropy</i> , (2018), 20(4):275. doi: 10.3390/e20040275.
[8]	Yu M, Bian X, Wang T, Wang J, Metal-based magnetic fluids with core-shell structure FeB@SiO <sub>2</sub> amorphous particles, <i>Soft Matter</i> , (2017), 13, 6340-6348; <a href="https://doi.org/10.1039/C7SM01238A">https://doi.org/10.1039/C7SM01238A</a>

Table A2: Results of the Chemdataextractor on the corpus. Note, we have highlighted in the text, the expected results for the NLP software, and then after the text, is the output from chemdataextractor, with each search term it found highlighted. Along with a note on how successful it was.

Paper		Result
[1]	<p>'Discovery of advanced soft-magnetic high entropy alloy (HEA) thin films are highly pursued to obtain unidentified functional materials. The figure of merit in current nanocrystalline HEA thin films relies in integration of a simple single-step electrochemical approach with a complex HEA system containing multiple elements with dissimilar crystal structures and large variation of melting points. A new family of <b>Cobalt-Copper-Iron-Nickel-Zinc (Co-Cu-Fe-Ni-Zn)</b> HEA thin films are prepared through pulse electrodeposition in aqueous medium, hosts nanocrystalline features in the range of ~5-20nm having FCC and BCC dual phases. <b>The fabricated Co-Cu-Fe-Ni-Zn HEA thin films exhibited high saturation magnetization value of ~82emu/g,</b> relatively <b>low coercivity value of 19.5 Oe</b> and remanent magnetization of 1.17%. Irrespective of the alloying of diamagnetic Zn and Cu with ferromagnetic Fe, Co, Ni elements, the HEA thin film has resulted in relatively high saturation magnetization which can provide useful insights for its potential unexplored applications.'</p> <div style="border: 1px solid black; padding: 10px; margin-top: 10px;"> <p><u>Output:</u></p> <pre>{'CoerciveField': {'raw_value': '19.5', 'raw_units': 'Oe', 'value': [19.5], 'units': 'Oersted^(1.0)', 'specifier': 'coercivity', 'compound': {'Compound': {'names': ['Co-Cu-Fe-Ni-Zn']}}}}</pre> <pre>{'MagnetizationMass': {'raw_value': '82', 'raw_units': 'emu/g', 'value': [82.0], 'units': 'Emu^(1.0) Gram^(-1.0)', 'specifier': 'saturation magnetization', 'compound': {'Compound': {'names': ['Co-Cu-Fe-Ni-Zn']}}}}</pre> <p>Success</p> </div>	Success
[2]	<p>'Multiprincipal-element alloys (MPEAs), including high-entropy alloys, are a new class of materials whose thermodynamical properties are mainly driven by configuration entropy, rather than enthalpy in the traditional alloys, especially at high temperatures. Herein, the design of a novel soft-magnetic nonequiatomic, quaternary MPEA is described, via tuning its chemical composition to deliberately manipulate its microstructure, such that it contains</p>	Success for all the properties and composition,

ultrafine ferromagnetic body-centered-cubic (BCC) coherent nanoprecipitates (3-7nm) uniformly distributed in a B2-phase matrix. The new alloy Al<sub>1.5</sub>Co<sub>4</sub>Fe<sub>2</sub>Cr exhibits high saturation magnetization ( $M_S = 135.3 \text{ emu g}^{-1}$ ), low coercivity ( $H_C = 127.3 \text{ A m}^{-1}$ ), high Curie temperature ( $T_C = 1061 \text{ K}$ ), and high electrical resistivity ( $\rho = 44 \mu\Omega \text{ cm}$ ), promising for soft magnets. More importantly, these prominent soft-magnetic properties are observed to be retained even after the alloy is thermally exposed at 873 K for 555 h, apparently attributable to the excellent stability of the coherent microstructure. The versatility of the magnetic properties of this new alloy is discussed in light of the microstructural change induced by tuning the chemical composition, and the enhanced performance of the alloy is compared directly with that of the traditional soft-magnetic alloys. The perspective is also addressed to design high-performance soft-magnetic alloys for high-temperature applications.'

but an additional magnetisation was also attributed for this alloy

#### Output:

```
['MagnetizationMass': {'raw_value': '135.3', 'raw_units': 'emu g-1', 'value': [135.3], 'units': 'Emu^(1.0) Gram^(-1.0)', 'specifier': 'saturation magnetization', 'compound': {'Compound': {'names': ['Al1.5 Co4 Fe2 Cr']}}}]
```

```
['CoerciveField': {'raw_value': '127.3', 'raw_units': 'Am-1', 'value': [127.3], 'units': 'Ampere^(1.0) Meter^(-1.0)', 'specifier': 'coercivity', 'compound': {'Compound': {'names': ['Al1.5 Co4 Fe2 Cr']}}}]
```

```
['CurieTemperature': {'raw_value': '1061', 'raw_units': 'K', 'value': [1061.0], 'units': 'Kelvin^(1.0)', 'specifier': 'Curie temperature TC', 'compound': {'Compound': {'names': ['Al1.5 Co4 Fe2 Cr']}}}]
```

#### Success

```
['MagnetizationVol': {'raw_value': '127.3', 'raw_units': 'Am-1', 'value': [127.3], 'units': 'Ampere^(1.0) Meter^(-1.0)', 'specifier': 'saturation magnetization', 'compound': {'Compound': {'names': ['Al1.5 Co4 Fe2 Cr']}}}]
```

Since  $\text{Am}^{-1}$  can also be a unit of magnetisation ChemDataExtractor has incorrectly attributed an additional magnetisation value for this alloy.

<p>[3]</p>	<p>'High silicon (Si) electrical steel has the potential for efficient use in applications such as electrical motors and generators with cost-effective in processing, but it is difficult to manufacture. Increasing the Si content beyond 3 wt.% improves magnetic and electrical properties, with 6.5 wt.% being achievable. The main goal of this research is to design, develop, and implement a scalable additive manufacturing process to fabricate Fe with 6.5 wt.% Si (Fe-6Si) steel with high magnetic permeability, high electrical resistivity, low coercivity, and low residual induction that other methods cannot achieve because of manufacturing limitations. Binder jet additive manufacturing was used to deposit near net shape components that were subsequently sintered via solid-state sintering to achieve near full densification. Here, it is shown that the use of solid-state sintering mitigates cracking since no rapid solidification occurs unlike fusion-based additive technologies. The Fe-6Si samples demonstrated an ultimate tensile strength of 434 MPa, electrical resistivity of 98 <math>\mu\Omega</math> cm, and saturation magnetization of 1.83 T with low coercivity and high permeability. The results strongly supports to replace the only available 0.1 mm thick chemical vapor deposition (CVD) produced Si steel using the cost effective AM method with good mechanical and magnetic properties for motor applications.'</p>	<p>Failed to extract a composition</p>
	<div style="border: 1px solid black; padding: 10px;"> <p><u>Output:</u></p> <pre>{'MagnetizationVol': {'raw_value': '1.83', 'raw_units': 'T', 'value': [1.83], 'units': 'Tesla^(1.0)', 'specifier': 'saturation magnetization', 'compound': {'Compound': {'names': [silicon]}}}}</pre> <p>Failed to identify Fe-6Si as a compound</p> </div>	
<p>[4]</p>	<p>'In this paper, a single phase <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder was successfully synthesized through the salt bath nitriding reaction method. The flaky FeSi alloy powder was used as the iron source, and non-toxic CO(NH<sub>2</sub>)<sub>2</sub> was used as the nitrogen source. The nitridation mechanism, the preparation technology, the soft magnetic properties, and the magnetization temperature dependence of the powder were studied. The research result showed that <math>\epsilon</math>-Fe(Si)<sub>3</sub>N alloy powders were synthesized in a high temperature nitrification system after the surface of flaky FeSi alloy powders were activated by a high-energy ball mill. The optimum nitriding process was nitridation for 1 h at 550 °C. The <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder had good thermal stability at less than 478.8 °C. It was shown that <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder has good soft magnetic properties, and the saturation magnetization of the powder was up to 139 emu/g. The saturation magnetization of <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder remains basically constant in the temperature range of 300-400 K. In the temperature range of 400-600 K, the saturation magnetization decreases slightly with the increase of temperature, indicating that the magnetic <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder has good magnetization temperature dependence.'</p>	<p>Composition recognised in-part.</p>

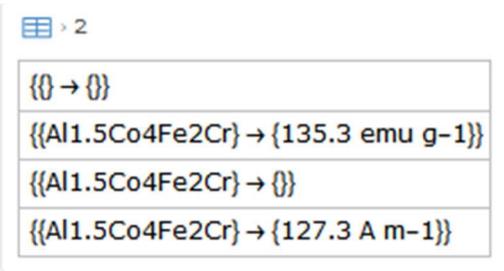
	<p><u>Output:</u></p> <pre>{'MagnetizationMass': {'raw_value': '139', 'raw_units': 'emu/g', 'value': [139.0], 'units': 'Emu^(1.0) Gram^(-1.0)', 'specifier': 'saturation magnetization', 'compound': {'Compound': {'names': ['FeSi']}}}}</pre> <p>Although it identified FeSi, it failed to identify <math>\epsilon</math>-Fe(Si)<sub>3</sub>N where the nitridation is key to the soft magnetic properties.</p>	
[5]	<p>'The soft magnetic properties of Fe-based nanocrystalline alloys are determined by their grain size. In the present article, the (Fe<sub>0.4</sub>Co<sub>0.6</sub>)<sub>79</sub>Nb<sub>3</sub>B<sub>18</sub> nanocrystalline alloys have been successfully prepared by isothermal annealing. The variation of soft magnetic properties as a function of annealing temperature and incubation time is investigated in detail. Two distinct crystallization behaviors were found for the (Fe<sub>0.4</sub>Co<sub>0.6</sub>)<sub>79</sub>Nb<sub>3</sub>B<sub>18</sub> alloys. The initial nanocrystallization products comprise a mixture of <math>\alpha</math>-Fe(Co), Fe<sub>2</sub>B, and Fe<sub>23</sub>B<sub>6</sub>-type crystalline metastable phases, and the final crystallization products are composed of <math>\alpha</math>-Fe(Co), Fe<sub>2</sub>B, and Fe<sub>3</sub>B crystalline phases. The grain size decreases first and then increases with the increasing annealing temperature in the range of 764~1151 K, and a fine grain size with mean grain size of 12.7 nm can be achieved for alloys annealed at 880 K. As the annealing temperature increases from 764 K to 1151 K, the saturation magnetization increases first and then decreases without a significant increase of the coercivity. The alloys annealed at 880 K exhibit the optimized soft magnetic properties with high Ms of 145 emu g<sup>-1</sup> and low Hc of 0.04 Oe.'</p> <p><u>Output:</u></p> <pre>{'MagnetizationMass': {'raw_value': '145', 'raw_units': 'emug-1', 'value': [145.0], 'units': 'Emu^(1.0) Gram^(-1.0)', 'specifier': 'Ms', 'compound': {'Compound': {'names': ['Fe']}}}}</pre> <pre>{'CoerciveField': {'raw_value': '0.04', 'raw_units': 'Oe', 'value': [0.04], 'units': 'Oersted^(1.0)', 'specifier': 'Hc', 'compound': {'Compound': {'names': ['FeZinc']}}}}</pre> <p>Failed to correctly identify (Fe<sub>0.4</sub>Co<sub>0.6</sub>)<sub>79</sub>Nb<sub>3</sub> as a compound.</p>	Failed to identify a composition

<p>[6]</p>	<p>FeSiBAlNi (W5), FeSiBAlNiCo (W6-Co), and FeSiBAlNiGd (W6-Gd) high entropy alloys (HEAs) were prepared using a copper-mold casting method. Effects of Co and Gd additions combined with subsequent annealing on microstructures and magnetism were investigated. The as-cast W5 consists of BCC solid solution and FeSi-rich phase. The Gd addition induces the formation of body-centered cubic (BCC) and face-centered cubic (FCC) solid solutions for W6-Gd HEAs. Whereas, the as-cast W6-Co is composed of the FeSi-rich phase. During annealing, no new phases arise in the W6-Co HEA, indicating a good phase stability. The as-cast W5 has the highest hardness (1210 HV), which is mainly attributed to the strengthening effect of FeSi-rich phase evenly distributed in the solid solution matrix. The tested FeSiBAlNi-based HEAs possess soft magnetism. The saturated magnetization and remanence ratio of W6-Gd are distinctly enhanced from 10.93 emu/g to 62.78 emu/g and from 1.44% to 15.50% after the annealing treatment, respectively. The good magnetism of the as-annealed W6-Gd can be ascribed to the formation of Gd-oxides.'</p>	<p>Failed to attribute the magnetisation to the correct composition</p>
	<div style="border: 1px solid black; padding: 10px;"> <p><u>Output:</u></p> <pre>{'MagnetizationMass': {'raw_value': '10.93 to 62.78', 'raw_units': 'emu/g', 'value': [10.93, 62.78], 'units': 'Emu^(1.0) Gram^(-1.0)', 'specifier': 'saturated magnetization', 'compound': {'Compound': {'names': ['FeSiBAlNi']}, 'labels': ['W5']}}}}</pre> <p>Failed to identify that this is the Gd substituted composition (W6-Gd) not the base FeSiBAlNi composition.</p> </div>	
<p>[7]</p>	<p>"Strong static magnetic field (SSMF) is a unique way to regulate the microstructure and improve the properties of materials. FeCoNi(AlSi)0.2 alloy is a novel class of soft magnetic materials (SMMs) designed based on high-entropy alloy (HEA) concepts. In this study, a strong static magnetic field is introduced to tune the microstructure, mechanical, electrical and magnetic properties of FeCoNi(AlSi)0.2 high-entropy alloy. Results indicate that, with the increasing magnetic field intensity, the Vickers hardness and the saturation magnetization (Ms) increase firstly, and then decrease and reach the maximum at 5T, while the yield strength, the residual magnetization (Mr) and the coercivity (Hc) take the opposite trend. The resistivity values (ρ) are found to be enhanced by the increasing magnetic field intensity. The main reasons for the magnetic field on the above effects are interpreted by microstructure evolution (phase species and volume fraction), atomic-level structure and defects (vacancy and dislocation density).'</p>	<p>Success</p>

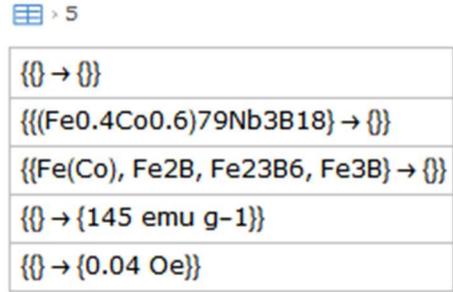
	<p><u>Output:</u></p> <pre>{'MagnetizationVol': {'raw_value': '5', 'raw_units': 'T', 'value': [5.0], 'units': 'Tesla^(1.0)', 'specifier': 'Ms', 'compound': {'Compound': {'names': ['FeCoNi(AlSi)0.2']}}}}</pre> <p>Success</p>	
[8]	<p>'FeB@SiO2 amorphous particles were firstly introduced into Ga85.8In14.2 alloys to prepare metal-based magnetic fluids. The morphology of the FeB amorphous particles is spherical with an average particle size of about 190 nm. The shape of the particles is regular and the particle size is homogeneous. Stable core-shell structure SiO2 modified FeB amorphous particles are obtained and the thickness of the SiO2 coatings is observed to be about 40 nm. The results of VSM confirm that the saturation magnetization of the FeB amorphous particles is 131.5 emu g<sup>-1</sup>, which is almost two times higher than that of the Fe3O4 particles. The saturation magnetization of the FeB@SiO2 amorphous particles is 106.9 emu g<sup>-1</sup>, an approximate decrease of 18.7% due to the non-magnetic SiO2 coatings. The results from the torsional oscillation viscometer show that the metal-based magnetic fluids with FeB amorphous particles exhibit a desirable high temperature performance and are ideal candidates for high temperature use.'</p> <p><u>Output:</u></p> <pre>{'MagnetizationMass': {'raw_value': '131.5', 'raw_units': 'emu g-1', 'value': [131.5], 'units': 'Emu^(1.0) Gram^(-1.0)', 'specifier': 'saturation magnetization', 'compound': {'Compound': {'names': ['FeB']}}}}</pre> <pre>{'MagnetizationMass': {'raw_value': '106.9', 'raw_units': 'emu g-1', 'value': [106.9], 'units': 'Emu^(1.0) Gram^(-1.0)', 'specifier': 'saturation magnetization', 'compound': {'Compound': {'names': ['FeB @ SiO2']}}}}</pre> <p>Success</p>	Success

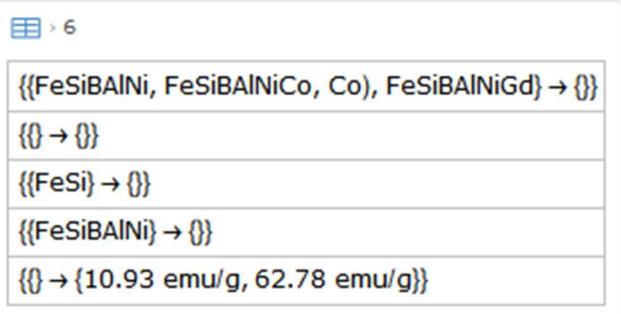
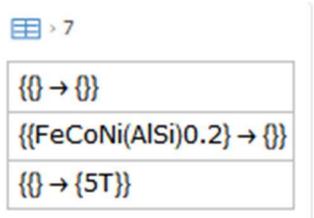
Table A3: Table A3: Results of the Regular Expression in combination with heuristics method on the corpus.

Paper		Result					
[1]	<p>'Discovery of advanced soft-magnetic high entropy alloy (HEA) thin films are highly pursued to obtain unidentified functional materials. The figure of merit in current nanocrystalline HEA thin films relies in integration of a simple single-step electrochemical approach with a complex HEA system containing multiple elements with dissimilar crystal structures and large variation of melting points. A new family of <b>Cobalt-Copper-Iron-Nickel-Zinc (Co-Cu-Fe-Ni-Zn)</b> HEA thin films are prepared through pulse electrodeposition in aqueous medium, hosts nanocrystalline features in the range of ~5-20nm having FCC and BCC dual phases. <b>The fabricated Co-Cu-Fe-Ni-Zn HEA thin films exhibited high saturation magnetization value of ~82emu/g,</b> relatively <b>low coercivity value of 19.5 Oe</b> and remanent magnetization of 1.17%. Irrespective of the alloying of diamagnetic Zn and Cu with ferromagnetic Fe, Co, Ni elements, the HEA thin film has resulted in relatively high saturation magnetization which can provide useful insights for its potential unexplored applications.'</p> <div data-bbox="331 708 786 1015" style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>☰ &gt; 1</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td style="padding: 2px;">{{{} → {}}}</td></tr> <tr><td style="padding: 2px;">{{{(Co-Cu-Fe-Ni-Zn)}} → {}}}</td></tr> <tr><td style="padding: 2px;">{{{Co-Cu-Fe-Ni-Zn}} → {82 emu/g}}</td></tr> <tr><td style="padding: 2px;">{{{Co-Cu-Fe-Ni-Zn}} → {19.5 Oe}}</td></tr> <tr><td style="padding: 2px;">{{{Co-Cu-Fe-Ni-Zn}} → {}}}</td></tr> </table> </div>	{{{} → {}}}	{{{(Co-Cu-Fe-Ni-Zn)}} → {}}}	{{{Co-Cu-Fe-Ni-Zn}} → {82 emu/g}}	{{{Co-Cu-Fe-Ni-Zn}} → {19.5 Oe}}	{{{Co-Cu-Fe-Ni-Zn}} → {}}}	Composition recognised, saturation magnetisation and coercivity picked up correctly,
{{{} → {}}}							
{{{(Co-Cu-Fe-Ni-Zn)}} → {}}}							
{{{Co-Cu-Fe-Ni-Zn}} → {82 emu/g}}							
{{{Co-Cu-Fe-Ni-Zn}} → {19.5 Oe}}							
{{{Co-Cu-Fe-Ni-Zn}} → {}}}							
[2]	<p>'Multiprincipal-element alloys (MPEAs), including high-entropy alloys, are a new class of materials whose thermodynamical properties are mainly driven by configuration entropy, rather than enthalpy in the traditional alloys, especially at high temperatures. Herein, the design of a novel soft-magnetic nonequiatomic, quaternary MPEA is described, via tuning its chemical composition to deliberately manipulate its microstructure, such that it contains ultrafine ferromagnetic body-centered-cubic (BCC) coherent nanoprecipitates (3-7nm) uniformly distributed in a B2-phase matrix. <b>The new alloy Al<sub>1.5</sub>Co<sub>4</sub>Fe<sub>2</sub>Cr exhibits high saturation magnetization (MS = 135.3 emu g<sup>-1</sup>), low coercivity (HC = 127.3 A m<sup>-1</sup>), high Curie temperature (TC = 1061 K),</b> and high electrical resistivity (ρ = 44 μΩ cm), promising for soft magnets. More importantly, these prominent soft-magnetic properties are observed to be retained even after the alloy is thermally exposed at 873 K for 555 h, apparently attributable to the excellent stability of the coherent microstructure. The versatility of the magnetic properties of this</p>	Composition recognised, saturation magnetisation and coercivity picked up correctly,					

	<p>new alloy is discussed in light of the microstructural change induced by tuning the chemical composition, and the enhanced performance of the alloy is compared directly with that of the traditional soft-magnetic alloys. The perspective is also addressed to design high-performance soft-magnetic alloys for high-temperature applications.'</p>									
	 <table border="1" data-bbox="324 343 828 614"> <tr> <td></td> <td></td> </tr> <tr> <td><math>\{\{\text{Al1.5Co4Fe2Cr}\} \rightarrow \{135.3 \text{ emu g}^{-1}\}\}</math></td> <td></td> </tr> <tr> <td><math>\{\{\text{Al1.5Co4Fe2Cr}\} \rightarrow \{\}\}</math></td> <td></td> </tr> <tr> <td><math>\{\{\text{Al1.5Co4Fe2Cr}\} \rightarrow \{127.3 \text{ A m}^{-1}\}\}</math></td> <td></td> </tr> </table>			$\{\{\text{Al1.5Co4Fe2Cr}\} \rightarrow \{135.3 \text{ emu g}^{-1}\}\}$		$\{\{\text{Al1.5Co4Fe2Cr}\} \rightarrow \{\}\}$		$\{\{\text{Al1.5Co4Fe2Cr}\} \rightarrow \{127.3 \text{ A m}^{-1}\}\}$		
$\{\{\text{Al1.5Co4Fe2Cr}\} \rightarrow \{135.3 \text{ emu g}^{-1}\}\}$										
$\{\{\text{Al1.5Co4Fe2Cr}\} \rightarrow \{\}\}$										
$\{\{\text{Al1.5Co4Fe2Cr}\} \rightarrow \{127.3 \text{ A m}^{-1}\}\}$										
[3]	<p>'High silicon (Si) electrical steel has the potential for efficient use in applications such as electrical motors and generators with cost-effective in processing, but it is difficult to manufacture. Increasing the Si content beyond 3 wt.% improves magnetic and electrical properties, with 6.5 wt.% being achievable. The main goal of this research is to design, develop, and implement a scalable additive manufacturing process to fabricate Fe with 6.5 wt.% Si (Fe-6Si) steel with high magnetic permeability, high electrical resistivity, low coercivity, and low residual induction that other methods cannot achieve because of manufacturing limitations. Binder jet additive manufacturing was used to deposit near net shape components that were subsequently sintered via solid-state sintering to achieve near full densification. Here, it is shown that the use of solid-state sintering mitigates cracking since no rapid solidification occurs unlike fusion-based additive technologies. The Fe-6Si samples demonstrated an ultimate tensile strength of 434 MPa, electrical resistivity of 98 <math>\mu\Omega</math> cm, and saturation magnetization of 1.83 T with low coercivity and high permeability. The results strongly supports to replace the only available 0.1 mm thick chemical vapor deposition (CVD) produced Si steel using the cost effective AM method with good mechanical and magnetic properties for motor applications.'</p>	<p>Composition recognised, saturation magnetisation picked up correctly</p>								

	<div data-bbox="331 220 591 517"> <p>☰ &gt; 3</p> <table border="1"> <tr><td>{{(Si)} → {}}</td></tr> <tr><td>{() → {}}</td></tr> <tr><td>{{(Fe-6Si)} → {}}</td></tr> <tr><td>{{(Fe-6Si)} → {83 T}}</td></tr> <tr><td>{{(Fe-6Si)} → {}}</td></tr> </table> </div>	{{(Si)} → {}}	{() → {}}	{{(Fe-6Si)} → {}}	{{(Fe-6Si)} → {83 T}}	{{(Fe-6Si)} → {}}	
{{(Si)} → {}}							
{() → {}}							
{{(Fe-6Si)} → {}}							
{{(Fe-6Si)} → {83 T}}							
{{(Fe-6Si)} → {}}							
[4]	<p>'In this paper, a single phase <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder was successfully synthesized through the salt bath nitriding reaction method. The flaky FeSi alloy powder was used as the iron source, and non-toxic CO(NH<sub>2</sub>)<sub>2</sub> was used as the nitrogen source. The nitridation mechanism, the preparation technology, the soft magnetic properties, and the magnetization temperature dependence of the powder were studied. The research result showed that <math>\epsilon</math>-Fe(Si)<sub>3</sub>N alloy powders were synthesized in a high temperature nitrification system after the surface of flaky FeSi alloy powders were activated by a high-energy ball mill. The optimum nitriding process was nitridation for 1 h at 550 °C. The <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder had good thermal stability at less than 478.8 °C. It was shown that <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder has good soft magnetic properties, and the saturation magnetization of the powder was up to 139 emu/g. The saturation magnetization of <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder remains basically constant in the temperature range of 300-400 K. In the temperature range of 400-600 K, the saturation magnetization decreases slightly with the increase of temperature, indicating that the magnetic <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder has good magnetization temperature dependence.'</p> <div data-bbox="331 957 685 1270"> <p>☰ &gt; 4</p> <table border="1"> <tr><td>{{(Fe(Si)3N)} → {}}</td></tr> <tr><td>{{(FeSi)} → {}}</td></tr> <tr><td>{() → {}}</td></tr> <tr><td>{{(Fe(Si)3N, FeSi)} → {}}</td></tr> <tr><td>{{(Fe(Si)3N)} → {139 emu/g}}</td></tr> </table> </div>	{{(Fe(Si)3N)} → {}}	{{(FeSi)} → {}}	{() → {}}	{{(Fe(Si)3N, FeSi)} → {}}	{{(Fe(Si)3N)} → {139 emu/g}}	Composition recognised, saturation magnetisation picked up correctly
{{(Fe(Si)3N)} → {}}							
{{(FeSi)} → {}}							
{() → {}}							
{{(Fe(Si)3N, FeSi)} → {}}							
{{(Fe(Si)3N)} → {139 emu/g}}							
[5]	<p>'The soft magnetic properties of Fe-based nanocrystalline alloys are determined by their grain size. In the present article, the (Fe<sub>0.4</sub>Co<sub>0.6</sub>)<sub>79</sub>Nb<sub>3</sub>B<sub>18</sub> nanocrystalline alloys have been successfully prepared by isothermal annealing. The variation of soft</p>	Composition recognised.					

	<p>magnetic properties as a function of annealing temperature and incubation time is investigated in detail. Two distinct crystallization behaviors were found for the (Fe<sub>0.4</sub>Co<sub>0.6</sub>)<sub>79</sub>Nb<sub>3</sub>B<sub>18</sub> alloys. The initial nanocrystallization products comprise a mixture of <math>\alpha</math>-Fe(Co), Fe<sub>2</sub>B, and Fe<sub>23</sub>B<sub>6</sub>-type crystalline metastable phases, and the final crystallization products are composed of <math>\alpha</math>-Fe(Co), Fe<sub>2</sub>B, and Fe<sub>3</sub>B crystalline phases. The grain size decreases first and then increases with the increasing annealing temperature in the range of 764~1151 K, and a fine grain size with mean grain size of 12.7 nm can be achieved for alloys annealed at 880 K. As the annealing temperature increases from 764 K to 1151 K, the saturation magnetization increases first and then decreases without a significant increase of the coercivity. The alloys annealed at 880 K exhibit the optimized soft magnetic properties with high Ms of 145 emu g<sup>-1</sup> and low Hc of 0.04 Oe.'</p>	<p>Values of saturation magnetisation and coercivity recognised, but no composition linked to these values. By elimination, it may be determined that it is (Fe<sub>0.4</sub>Co<sub>0.6</sub>)<sub>79</sub>Nb<sub>3</sub>B<sub>18</sub></p>
	 <p>The screenshot shows a search results window with a list of five items:</p> <ul style="list-style-type: none"> <li>{ } → { }</li> <li>{(Fe<sub>0.4</sub>Co<sub>0.6</sub>)<sub>79</sub>Nb<sub>3</sub>B<sub>18</sub>} → { }</li> <li>{(Fe(Co), Fe<sub>2</sub>B, Fe<sub>23</sub>B<sub>6</sub>, Fe<sub>3</sub>B)} → { }</li> <li>{ } → {145 emu g<sup>-1</sup>}</li> <li>{ } → {0.04 Oe}}</li> </ul>	
[6]	<p>FeSiBAlNi (W5), FeSiBAlNiCo (W6-Co), and FeSiBAlNiGd (W6-Gd) high entropy alloys (HEAs) were prepared using a copper-mold casting method. Effects of Co and Gd additions combined with subsequent annealing on microstructures and magnetism were investigated. The as-cast W5 consists of BCC solid solution and FeSi-rich phase. The Gd addition induces the formation of body-centered cubic (BCC) and face-centered cubic (FCC) solid solutions for W6-Gd HEAs. Whereas, the as-cast W6-Co is composed of the FeSi-rich phase. During annealing, no new phases arise in the W6-Co HEA, indicating a good phase stability. The as-cast W5 has the highest hardness (1210 HV), which is mainly attributed to the strengthening effect of FeSi-rich phase evenly distributed in the solid solution matrix. The tested FeSiBAlNi-based HEAs possess soft magnetism. The saturated magnetization and remanence ratio of W6-Gd are distinctly enhanced from 10.93 emu/g to 62.78 emu/g and from 1.44% to 15.50% after the annealing treatment, respectively. The good magnetism of the as-annealed W6-Gd can be ascribed to the formation of Gd-oxides.'</p>	<p>Composition recognised. Values of saturation magnetisation and coercivity recognised, but no composition linked to these values. This is</p>

		because of abbreviations being used.
[7]	<p>“Strong static magnetic field (SSMF) is a unique way to regulate the microstructure and improve the properties of materials. <b>FeCoNi(AlSi)0.2</b> alloy is a novel class of soft magnetic materials (SMMs) designed based on high-entropy alloy (HEA) concepts. In this study, a strong static magnetic field is introduced to tune the microstructure, mechanical, electrical and magnetic properties of FeCoNi(AlSi)0.2 high-entropy alloy. Results indicate that, with the increasing magnetic field intensity, the Vickers hardness and the <b>saturation magnetization (Ms) increase firstly, and then decrease and reach the maximum at 5T</b>, while the yield strength, the residual magnetization (Mr) and the coercivity (Hc) take the opposite trend. The resistivity values (<math>\rho</math>) are found to be enhanced by the increasing magnetic field intensity. The main reasons for the magnetic field on the above effects are interpreted by microstructure evolution (phase species and volume fraction), atomic-level structure and defects (vacancy and dislocation density).’</p> 	Composition recognised. Values of saturation magnetisation recognised, but no composition linked. By elimination, it may be determined that it is <b>FeCoNi(AlSi)0.2</b>
[8]	<p><b>'FeB@SiO2 amorphous particles were firstly introduced into Ga85.8In14.2 alloys to prepare metal-based magnetic fluids.</b> The morphology of the FeB amorphous particles is spherical with an average particle size of about 190 nm. The shape of the particles is regular and the particle size is homogeneous. Stable core-shell structure SiO2 modified FeB amorphous particles are obtained and the thickness of the SiO2 coatings is observed to be about 40 nm. The results of VSM confirm that <b>the saturation magnetization of the FeB amorphous particles is 131.5 emu g<sup>-1</sup></b>, which is almost two times higher than that of the Fe3O4 particles. <b>The saturation magnetization of the FeB@SiO2 amorphous particles is 106.9 emu g<sup>-1</sup></b>, an</p>	Composition recognised, saturation magnetisation picked up correctly

approximate decrease of 18.7% due to the non-magnetic SiO<sub>2</sub> coatings. The results from the torsional oscillation viscometer show that the metal-based magnetic fluids with FeB amorphous particles exhibit a desirable high temperature performance and are ideal candidates for high temperature use.'

☰ > 8

{{FeB, SiO<sub>2</sub>, Ga<sub>85.8</sub>In<sub>14.2</sub>} → {}}

{{FeB} → {}}

{{} → {}}

{{SiO<sub>2</sub>, FeB} → {}}

{{FeB, Fe<sub>3</sub>O<sub>4</sub>} → {131.5 emu g<sup>-1</sup>}}

{{FeB, Fe<sub>3</sub>O<sub>4</sub>} → {}}

{{FeB, SiO<sub>2</sub>} → {106.9 emu g<sup>-1</sup>}}

{{FeB, SiO<sub>2</sub>} → {}}

**Table A4: Results of the MatSciBERT software on the corpus.**

As MatSciBERT gives a label to every word with the abstract, this means that the output for the data is not as simple as the other two data sets. Rather the raw data output is 100's of lines long. Thus the raw data output for the first corpus is found here: [here](#)

The analysed data is given in the table below.

Paper		Result
[1]	<p>'Discovery of advanced soft-magnetic high entropy alloy (HEA) thin films are highly pursued to obtain unidentified functional materials. The figure of merit in current nanocrystalline HEA thin films relies in integration of a simple single-step electrochemical approach with a complex HEA system containing multiple elements with dissimilar crystal structures and large variation of melting points. A new family of <b>Cobalt-Copper-Iron-Nickel-Zinc (Co-Cu-Fe-Ni-Zn)</b> HEA thin films are prepared through pulse electrodeposition in aqueous medium, hosts nanocrystalline features in the range of ~5-20nm having FCC and BCC dual phases. <b>The fabricated Co-Cu-Fe-Ni-Zn HEA thin films exhibited high saturation magnetization value of ~82emu/g, relatively low coercivity value of 19.5 Oe</b> and remanent magnetization of 1.17%. Irrespective of the alloying of diamagnetic Zn and Cu with ferromagnetic Fe, Co, Ni elements, the HEA thin film has resulted in relatively high saturation magnetization which can provide useful insights for its potential unexplored applications.'</p> <pre> discovery of advanced soft - magnetic high entropy alloy ( hea ) thin films are highly pursued to obtain unidentified functional materials . the figure of merit in current nanoc rystalline hea thin films relies in integration of a simple single - step electrochemical approach with a complex hea system containing multiple elements with dissimilar crystal structures and large variation of melting points . a new family of cobalt - copper - iron - nickel - zinc ( co - cu - fe - ni - zn ) hea thin films are prepared through pulse electrode position in aqueous medium hosts nanoc rystalline features in the range of ~ 5 - 20 nm having fcc and bcc dual phases . the fabricated co - cu - fe - ni - zn hea thin films exhibited high saturation magnetization value of ~ 82 em u / g relatively low coerc ivity value of 19 . 5 oe and rem anen t magnetization of 1 . 17 % . irrespective of the alloy ing of diam agnetic zn and cu with ferromagnetic fe co ni elements the hea thin film has resulted in relatively high saturation magnetization which can provide useful insights for its potential unexpl ored applications .                     </pre>	None

[2]	<p>'Multiprincipal-element alloys (MPEAs), including high-entropy alloys, are a new class of materials whose thermodynamical properties are mainly driven by configuration entropy, rather than enthalpy in the traditional alloys, especially at high temperatures. Herein, the design of a novel soft-magnetic nonequiatomic, quaternary MPEA is described, via tuning its chemical composition to deliberately manipulate its microstructure, such that it contains ultrafine ferromagnetic body-centered-cubic (BCC) coherent nanoprecipitates (3-7nm) uniformly distributed in a B2-phase matrix. The new alloy Al<sub>1.5</sub>Co<sub>4</sub>Fe<sub>2</sub>Cr exhibits high saturation magnetization (<math>M_S = 135.3 \text{ emu g}^{-1}</math>), low coercivity (<math>H_C = 127.3 \text{ A m}^{-1}</math>), high Curie temperature (<math>T_C = 1061 \text{ K}</math>), and high electrical resistivity (<math>\rho = 44 \text{ } \mu\Omega \text{ cm}</math>), promising for soft magnets. More importantly, these prominent soft-magnetic properties are observed to be retained even after the alloy is thermally exposed at 873 K for 555 h, apparently attributable to the excellent stability of the coherent microstructure. The versatility of the magnetic properties of this new alloy is discussed in light of the microstructural change induced by tuning the chemical composition, and the enhanced performance of the alloy is compared directly with that of the traditional soft-magnetic alloys. The perspective is also addressed to design high-performance soft-magnetic alloys for high-temperature applications.'</p>	None
	<p>multi principal - element alloys ( mpeas ) including high - entropy alloys are a new class of materials whose thermodynamic al properties are mainly driven by configuration entropy rather than enthalpy in the traditional alloys especially at high temperatures . herein the design of a novel soft - magnetic none qu ia tom ic quaternary mpea is described via tuning its chemical composition to deliberately manipulate its microstructure such that it contains ultraf ine ferromagnetic body - centered - cubic ( bcc ) coherent nanop recip itates ( 3 - 7 nm ) uniformly distributed in a b 2 - phase matrix . the new alloy al 1 . 5 co 4 fe 2 cr exhibits high saturation magnetization ( ms = 135 . 3 em u g - 1 ) low coerc ivity ( hc = 127 . 3 a m - 1 ) high cur ie temperature ( tc = 106 1 k ) and high electrical resistivity ( <math>\rho = 44 \mu \omega \text{ cm}</math> ) promising for soft magnets . more importantly these prominent soft - magnetic properties are observed to be retained even after the alloy is thermally exposed at 87 3 k for 55 5 h apparently attributable to the excellent stability of the coherent microstructure . the versa tility of the magnetic properties of this new alloy is discussed in light of the micros tructural change induced by tuning the chemical composition and the enhanced performance of the alloy is compared directly with that of the traditional soft - magnetic alloys . the perspective is also addressed to design high - performance soft - magnetic alloys for high - temperature applications .</p>	

[3]	<p>'High silicon (Si) electrical steel has the potential for efficient use in applications such as electrical motors and generators with cost-effective in processing, but it is difficult to manufacture. Increasing the Si content beyond 3 wt.% improves magnetic and electrical properties, with 6.5 wt.% being achievable. The main goal of this research is to design, develop, and implement a scalable additive manufacturing process to fabricate Fe with 6.5 wt.% Si (Fe-6Si) steel with high magnetic permeability, high electrical resistivity, low coercivity, and low residual induction that other methods cannot achieve because of manufacturing limitations. Binder jet additive manufacturing was used to deposit near net shape components that were subsequently sintered via solid-state sintering to achieve near full densification. Here, it is shown that the use of solid-state sintering mitigates cracking since no rapid solidification occurs unlike fusion-based additive technologies. The Fe-6Si samples demonstrated an ultimate tensile strength of 434 MPa, electrical resistivity of 98 <math>\mu\Omega</math> cm, and saturation magnetization of 1.83 T with low coercivity and high permeability. The results strongly supports to replace the only available 0.1 mm thick chemical vapor deposition (CVD) produced Si steel using the cost effective AM method with good mechanical and magnetic properties for motor applications.'</p>	None
	<p>high silicon ( si ) electrical steel has the potential for efficient use in applications such as electrical motors and generators with cost - effective in processing but it is difficult to manufacture . increasing the si content beyond 3 wt . % improves magnetic and electrical properties with 6 . 5 wt . % being achievable . the main goal of this research is to design develop and implement a scalable additive manufacturing process to fabric ate fe with 6 . 5 wt . % si ( fe - 6 si ) steel with high magnetic permeability high electrical resistivity low coerc ivity and low residual induction that other methods cannot achieve because of manufacturing limitations . binder jet additive manufacturing was used to deposit near net shape components that were subsequently sin tered via solid - state sin tering to achieve near full dens ification . here it is shown that the use of solid - state sin tering mitigate s crack ing since no rapid solid ification occurs unlike fusion - based additive technologies . the fe - 6 si samples demonstrated an ultimate tensile strength of 43 4 mpa electrical resistivity of 98 <math>\mu \omega</math> cm and saturation magnetization of 1 . 83 t with low coerc ivity and high permeability . the results strongly supports to replace the only available 0 . 1 mm thick chemical vapor deposition ( cvd ) produced si steel using the cost effective am method with good mechanical and magnetic properties for motor applications .</p>	
[4]	<p>'In this paper, a single phase <math>\epsilon</math>-Fe(Si)<sub>3</sub>N powder was successfully synthesized through the salt bath nitriding reaction method. The flaky</p>	None

	<p>FeSi alloy powder was used as the iron source, and non-toxic <math>\text{CO}(\text{NH}_2)_2</math> was used as the nitrogen source. The nitridation mechanism, the preparation technology, the soft magnetic properties, and the magnetization temperature dependence of the powder were studied. The research result showed that <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> alloy powders were synthesized in a high temperature nitrification system after the surface of flaky FeSi alloy powders were activated by a high-energy ball mill. The optimum nitriding process was nitridation for 1 h at 550 °C. The <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> powder had good thermal stability at less than 478.8 °C. It was shown that <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> powder has good soft magnetic properties, and the saturation magnetization of the powder was up to 139 emu/g. The saturation magnetization of <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> powder remains basically constant in the temperature range of 300-400 K. In the temperature range of 400-600 K, the saturation magnetization decreases slightly with the increase of temperature, indicating that the magnetic <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> powder has good magnetization temperature dependence.'</p>	
	<p>in this paper a single phase <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> powder was successfully synthesized through the salt bath nitriding reaction method. the flaky FeSi alloy powder was used as the iron source and non-toxic <math>\text{CO}(\text{NH}_2)_2</math> was used as the nitrogen source. the nitridation mechanism the preparation technology the soft magnetic properties and the magnetization temperature dependence of the powder were studied. the research result showed that <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> alloy powders were synthesized in a high temperature nitrification system after the surface of flaky FeSi alloy powders were activated by a high-energy ball mill. the optimum nitriding process was nitridation for 1 h at 550 °C. the <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> powder had good thermal stability at less than 478.8 °C. it was shown that <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> powder has good soft magnetic properties and the saturation magnetization of the powder was up to 139 emu/g. the saturation magnetization of <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> powder remains basically constant in the temperature range of 300 - 400 K. in the temperature range of 400 - 600 K the saturation magnetization decreases slightly with the increase of temperature indicating that the magnetic <math>\epsilon\text{-Fe}(\text{Si})_3\text{N}</math> powder has good magnetization temperature dependence. view full-text</p>	
[5]	<p>'The soft magnetic properties of Fe-based nanocrystalline alloys are determined by their grain size. In the present article, the <math>(\text{Fe}_{0.4}\text{Co}_{0.6})_{79}\text{Nb}_3\text{B}_{18}</math> nanocrystalline alloys have been successfully prepared by isothermal annealing. The variation of soft magnetic properties as a function of annealing temperature and incubation time is investigated in detail. Two distinct crystallization behaviors were found for the <math>(\text{Fe}_{0.4}\text{Co}_{0.6})_{79}\text{Nb}_3\text{B}_{18}</math> alloys. The initial nanocrystallization products comprise a mixture of <math>\alpha\text{-Fe}(\text{Co})</math>, <math>\text{Fe}_2\text{B}</math>, and <math>\text{Fe}_{23}\text{B}_6</math>-type crystalline metastable phases, and the final crystallization products are composed of <math>\alpha\text{-Fe}(\text{Co})</math>, <math>\text{Fe}_2\text{B}</math>, and <math>\text{Fe}_3\text{B}</math> crystalline phases. The grain size decreases first and then increases with the increasing annealing temperature in the range of 764-1151 K, and a fine grain size with</p>	None

	<p>mean grain size of 12.7 nm can be achieved for alloys annealed at 880 K. As the annealing temperature increases from 764 K to 1151 K, the saturation magnetization increases first and then decreases without a significant increase of the coercivity. The alloys annealed at 880 K exhibit the optimized soft magnetic properties with high Ms of 145 emu g<sup>-1</sup> and low Hc of 0.04 Oe.'</p>	
	<p>the soft magnetic properties of fe - based nanoc rystalline alloys are determined by their grain size . in the present article the ( fe 0 . 4 co 0 . 6 ) 79 nb 3 b 18 nanoc rystalline alloys have been successfully prepared by isothermal annealing . the variation of soft magnetic properties as a function of annealing temperature and incubation time is investigated in detail . two distinct crystallization behaviors were found for the ( fe 0 . 4 co 0 . 6 ) 79 nb 3 b 18 alloys . the initial nanoc rystall ization products comprise a mixture of α - fe ( co ) fe 2 b and fe 23 b 6 - type crystalline metast able phases and the final crystallization products are composed of α - fe ( co ) fe 2 b and fe 3 b crystalline phases . the grain size decreases first and then increases with the increasing annealing temperature in the range of 76 4 - 115 1 k and a fine grain size with mean grain size of 12 . 7 nm can be achieved for alloys annealed at 88 0 k . as the annealing temperature increases from 76 4 k to 115 1 k the saturation magnetization increases first and then decreases without a significant increase of the coerc ivity . the alloys annealed at 88 0 k exhibit the optimized soft magnetic properties with high ms of 145 em u g -1 and low hc of 0 . 04 oe . view full - text</p>	
[6]	<p>FeSiBAlNi (W5), FeSiBAlNiCo (W6-Co), and FeSiBAlNiGd (W6-Gd) high entropy alloys (HEAs) were prepared using a copper-mold casting method. Effects of Co and Gd additions combined with subsequent annealing on microstructures and magnetism were investigated. The as-cast W5 consists of BCC solid solution and FeSi-rich phase. The Gd addition induces the formation of body-centered cubic (BCC) and face-centered cubic (FCC) solid solutions for W6-Gd HEAs. Whereas, the as-cast W6-Co is composed of the FeSi-rich phase. During annealing, no new phases arise in the W6-Co HEA, indicating a good phase stability. The as-cast W5 has the highest hardness (1210 HV), which is mainly attributed to the strengthening effect of FeSi-rich phase evenly distributed in the solid solution matrix. The tested FeSiBAlNi-based HEAs possess soft magnetism. The saturated magnetization and remanence ratio of W6-Gd are distinctly enhanced from 10.93 emu/g to 62.78 emu/g and from 1.44% to 15.50% after the annealing treatment, respectively. The good magnetism of the as-annealed W6-Gd can be ascribed to the formation of Gd-oxides.'</p>	None

	<p> <math>\text{Fe}_{50}\text{Co}_{50}</math> (<math>w_5</math>) <math>\text{Fe}_{50}\text{Co}_{40}\text{Ni}_{10}</math> (<math>w_6 - \text{Co}</math>) and <math>\text{Fe}_{50}\text{Co}_{40}\text{Ni}_{10}\text{Gd}_{10}</math> (<math>w_6 - \text{Gd}</math>) high entropy alloys (HEAs) were prepared using a copper - mold casting method . effects of <math>\text{Co}</math> and <math>\text{Gd}</math> additions combined with subsequent annealing on microstructures and magnetism were investigated . the as - cast <math>w_5</math> consists of bcc solid solution and <math>\text{Fe}_{50}\text{Co}_{50}</math> - rich phase . the <math>\text{Gd}</math> addition induces the formation of body - centered cubic ( bcc ) and face - centered cubic ( fcc ) solid solutions for <math>w_6 - \text{Gd}</math> HEAs . whereas the as - cast <math>w_6 - \text{Co}</math> is composed of the <math>\text{Fe}_{50}\text{Co}_{50}</math> - rich phase . during annealing no new phases arise in the <math>w_6 - \text{Co}</math> HEA indicating a good phase stability . the as - cast <math>w_5</math> has the highest hardness ( 1210 HV ) which is mainly attributed to the strengthening effect of <math>\text{Fe}_{50}\text{Co}_{50}</math> - rich phase evenly distributed in the solid solution matrix . the tested <math>\text{Fe}_{50}\text{Co}_{50}\text{Ni}_{10}</math> - based HEAs possess soft magnetism . the saturated magnetization and remanence ratio of <math>w_6 - \text{Gd}</math> are distinctly enhanced from 10 . 93 emu / g to 62 . 78 emu / g and from 1 . 44 % to 15 . 50 % after the annealing treatment respectively . the good magnetism of the as - annealed <math>w_6 - \text{Gd}</math> can be ascribed to the formation of <math>\text{Gd}</math> - oxides . view full - text </p>	
[7]	<p> “Strong static magnetic field (SSMF) is a unique way to regulate the microstructure and improve the properties of materials. <b>FeCoNi(AlSi)0.2</b> alloy is a novel class of soft magnetic materials (SMMs) designed based on high-entropy alloy (HEA) concepts. In this study, a strong static magnetic field is introduced to tune the microstructure, mechanical, electrical and magnetic properties of FeCoNi(AlSi)0.2 high-entropy alloy. Results indicate that, with the increasing magnetic field intensity, the Vickers hardness and the <b>saturation magnetization (<math>M_s</math>) increase firstly, and then decrease and reach the maximum at 5T</b>, while the yield strength, the residual magnetization (<math>M_r</math>) and the coercivity (<math>H_c</math>) take the opposite trend. The resistivity values (<math>\rho</math>) are found to be enhanced by the increasing magnetic field intensity. The main reasons for the magnetic field on the above effects are interpreted by microstructure evolution (phase species and volume fraction), atomic-level structure and defects (vacancy and dislocation density).” </p>	None

	<p>strong static magnetic field ( ss mf ) is a unique way to regulate the microstructure and improve the properties of materials . fec oni ( als i ) 0 . 2 alloy is a novel class of soft magnetic materials ( sm ms ) designed based on high - entropy alloy ( hea ) concepts . in this study a strong static magnetic field is introduced to tune the microstructure mechanical electrical and magnetic properties of fec oni ( als i ) 0 . 2 high - entropy alloy . results indicate that with the increasing magnetic field intensity the vic kers hardness and the saturation magnetization ( ms ) increase firstly and then decrease and reach the maximum at 5 t while the yield strength the residual magnetization ( mr ) and the coerc ivity ( hc ) take the opposite trend . the resistivity values ( ρ ) are found to be enhanced by the increasing magnetic field intensity . the main reasons for the magnetic field on the above effects are interpreted by microstructure evolution ( phase species and volume fraction ) atomic - level structure and defects ( vac ancy and dislocation density ) . view full - text</p>	
[8]	<p>'FeB@SiO<sub>2</sub> amorphous particles were firstly introduced into Ga<sub>85.8</sub>In<sub>14.2</sub> alloys to prepare metal-based magnetic fluids. The morphology of the FeB amorphous particles is spherical with an average particle size of about 190 nm. The shape of the particles is regular and the particle size is homogeneous. Stable core-shell structure SiO<sub>2</sub> modified FeB amorphous particles are obtained and the thickness of the SiO<sub>2</sub> coatings is observed to be about 40 nm. The results of VSM confirm that the saturation magnetization of the FeB amorphous particles is 131.5 emu g<sup>-1</sup>, which is almost two times higher than that of the Fe<sub>3</sub>O<sub>4</sub> particles. The saturation magnetization of the FeB@SiO<sub>2</sub> amorphous particles is 106.9 emu g<sup>-1</sup>, an approximate decrease of 18.7% due to the non-magnetic SiO<sub>2</sub> coatings. The results from the torsional oscillation viscometer show that the metal-based magnetic fluids with FeB amorphous particles exhibit a desirable high temperature performance and are ideal candidates for high temperature use.'</p>	None

Fe<sub>3</sub>O<sub>4</sub>@SiO<sub>2</sub> amorphous particles were firstly introduced into Ga<sub>85</sub>.8In<sub>14</sub>.2 alloys to prepare metal - based magnetic fluids . the morphology of the Fe<sub>3</sub>O<sub>4</sub> amorphous particles is spherical with an average particle size of about 190 nm . the shape of the particles is regular and the particle size is homogeneous . stable core - shell structure SiO<sub>2</sub> modified Fe<sub>3</sub>O<sub>4</sub> amorphous particles are obtained and the thickness of the SiO<sub>2</sub> coatings is observed to be about 40 nm . the results of vs m confirm that the saturation magnetization of the Fe<sub>3</sub>O<sub>4</sub> amorphous particles is 131 . 5 em u g<sup>-1</sup> which is almost two times higher than that of the Fe<sub>3</sub>O<sub>4</sub> particles . the saturation magnetization of the Fe<sub>3</sub>O<sub>4</sub>@SiO<sub>2</sub> amorphous particles is 106 . 9 em u g<sup>-1</sup> an approximate decrease of 18 . 7 % due to the non - magnetic SiO<sub>2</sub> coatings . the results from the torsion al oscillation visco meter show that the metal - based magnetic fluids with Fe<sub>3</sub>O<sub>4</sub> amorphous particles exhibit a desirable high temperature performance and are ideal candidates for high temperature use .

Figure A1: Screenshots of the Jupyter notebook used during the NLP part of the project.

## Creating model for extracting magnetisation properties using ChemDataExtractor

Import relevant methods from ChemDataExtractor

```
[29]: #!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Mon Dec 13 16:19:08 2021

@author: richard
"""
from chemdataextractor.doc import Document, Paragraph, Sentence
from chemdataextractor.model.model import Compound, ModelType, StringType
from chemdataextractor.parse.elements import I, R, T
from chemdataextractor.parse.actions import join
from chemdataextractor.model.units import TemperatureModel, MagneticFieldModel, MagnetizationByVolModel, MagnetizationByMas:

import numpy as np
import json
```

We need to define a new model since magnetisation parameter extraction isn't natively included in ChemDataExtractor.

We will start with magnetisation by volume. For this we will need to define a specifier expression typically seen in publications for this parameter. Here we use the regular expression 'saturat(?:ed|ion) + magneti[sz]ation' or 'Ms'.

We will also need to define a compound, so that ChemDataExtractor looks for a compound tied to the magnetisation by volume parameter it finds in text. We can use the existing tools for this as ChemDataExtractor already deals with a variety of compounds.

compound and specifier are keywords inherited from the MagnetizationByVolModel class in the argument. The MagnetizationByVolModel was generated as an additional model within the ChemDataExtractor software. This is straightforward but cumbersome, as it is required for every parameter with new units. It is done by updating the QuantityModel class as outlined here: [https://cambridgemolecularengineering-chemdataextractor-development.readthedocs-hosted.com/en/latest/migration\\_guide.html](https://cambridgemolecularengineering-chemdataextractor-development.readthedocs-hosted.com/en/latest/migration_guide.html)

```
[30]: class MagnetizationVol(MagnetizationByVolModel):
    specifier_expr = ((R('saturat(?:ed|ion)') + R('magneti[sz]ation')) | R('Ms')).add_action(join)
    specifier = StringType(parse_expression=specifier_expr,
                           required=True, contextual=True, updatable=True)
    compound = ModelType(Compound, required=True, contextual=True)
```

We do the same for all other magnetic properties we are interested in extracting. A before, anything with new units e.g. magnetisation by mass and magnetic field requires a new model to be made within the ChemDataExtractor code itself. Note that there already existed a TemperatureModel in ChemDataExtractor so there was no need to create one for the CurieTemperature model. Likewise, if I wanted to extract the saturation magnetic field I could reuse the MagneticFieldModel I created for the CoerciveField extraction mode.

```
[31]: class MagnetizationMass(MagnetizationByMassModel):
    specifier_expr = ((R('saturat(?:ed|ion)') + R('magneti[sz]ation')) | R('Ms')).add_action(join)
    specifier = StringType(parse_expression=specifier_expr,
                           required=True, contextual=True, updatable=True)
    compound = ModelType(Compound, required=True, contextual=True, updatable=False)

class CoerciveField(MagneticFieldModel):
    specifier_expr = ((I('coercive') + I('field')) | I('coercivity') | R('Hc')).add_action(join)
    specifier = StringType(parse_expression=specifier_expr,
                           required=True, contextual=True, updatable=True)
    compound = ModelType(Compound, required=True, contextual=True, updatable=False)

class CurieTemperature(TemperatureModel):
    specifier_expression = ((I('Curie') + I('temperature')) | I('Tc')).add_action(join)
    specifier = StringType(parse_expression=specifier_expression, required=True,
                           contextual=True, updatable=True)
    compound = ModelType(Compound, required=True, contextual=True, updatable=False)
```



Figure A3: Schematic of the high through-put experimental procedure

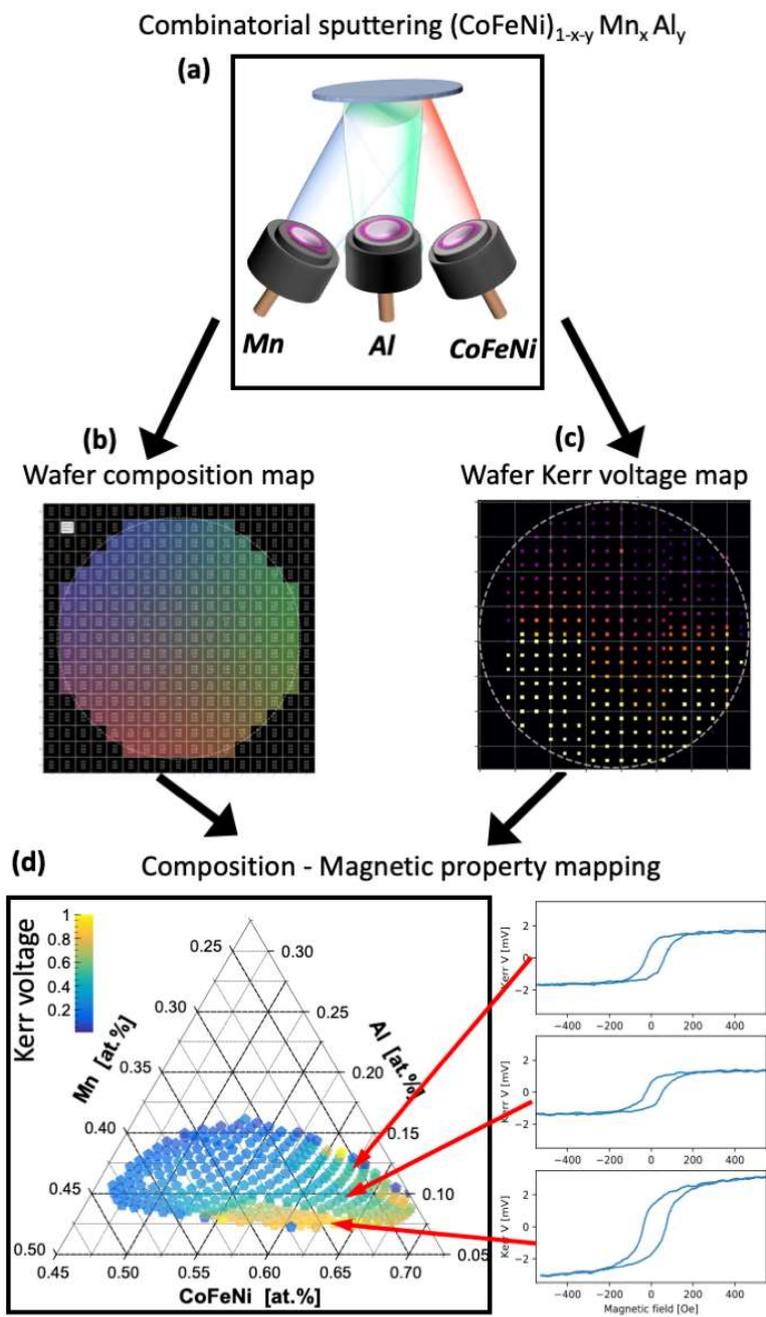


Figure A4: Photograph of the film with the coordinate system.

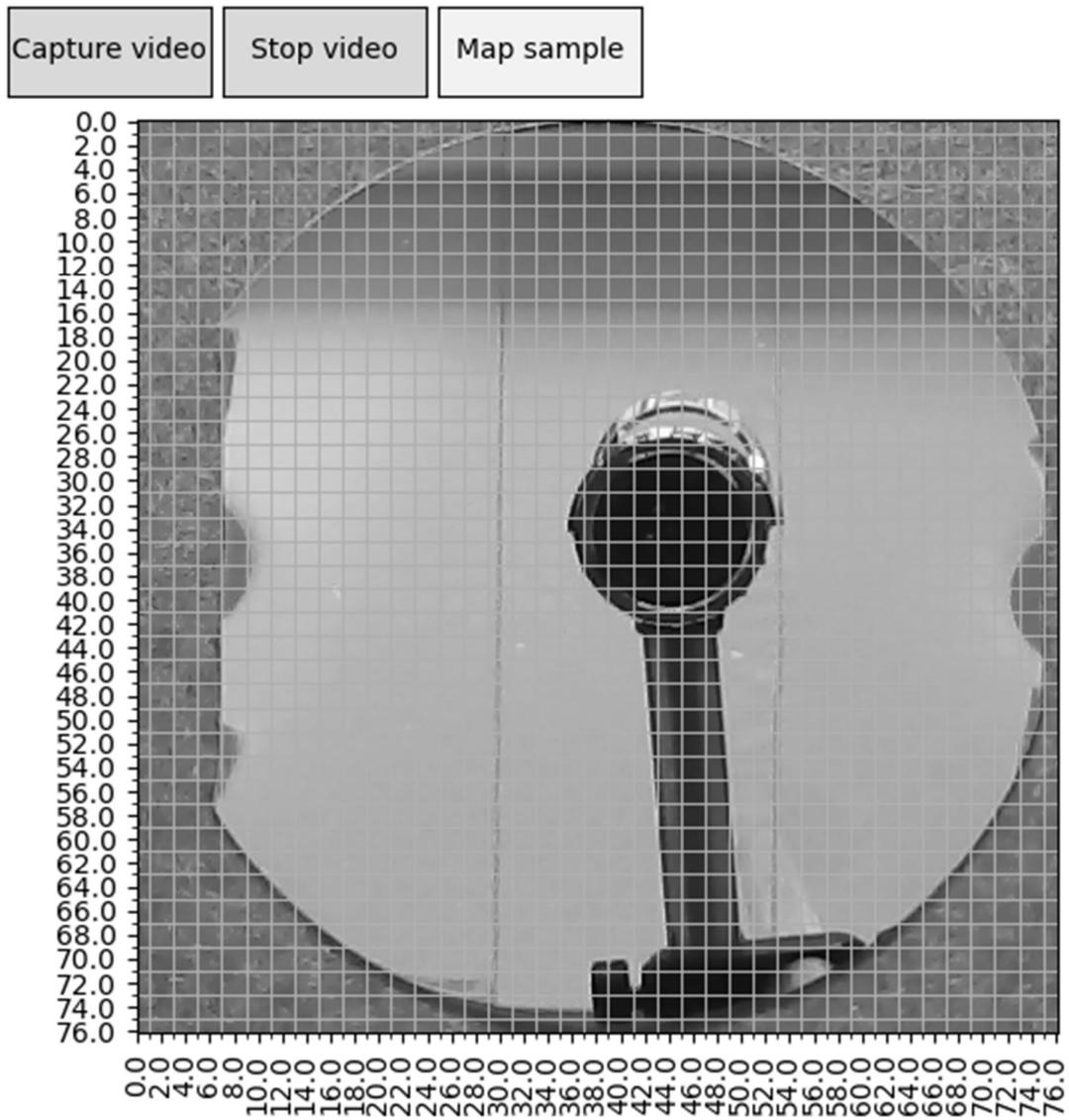


Figure A5: MOKE hysteresis loops measured on the high through-put system

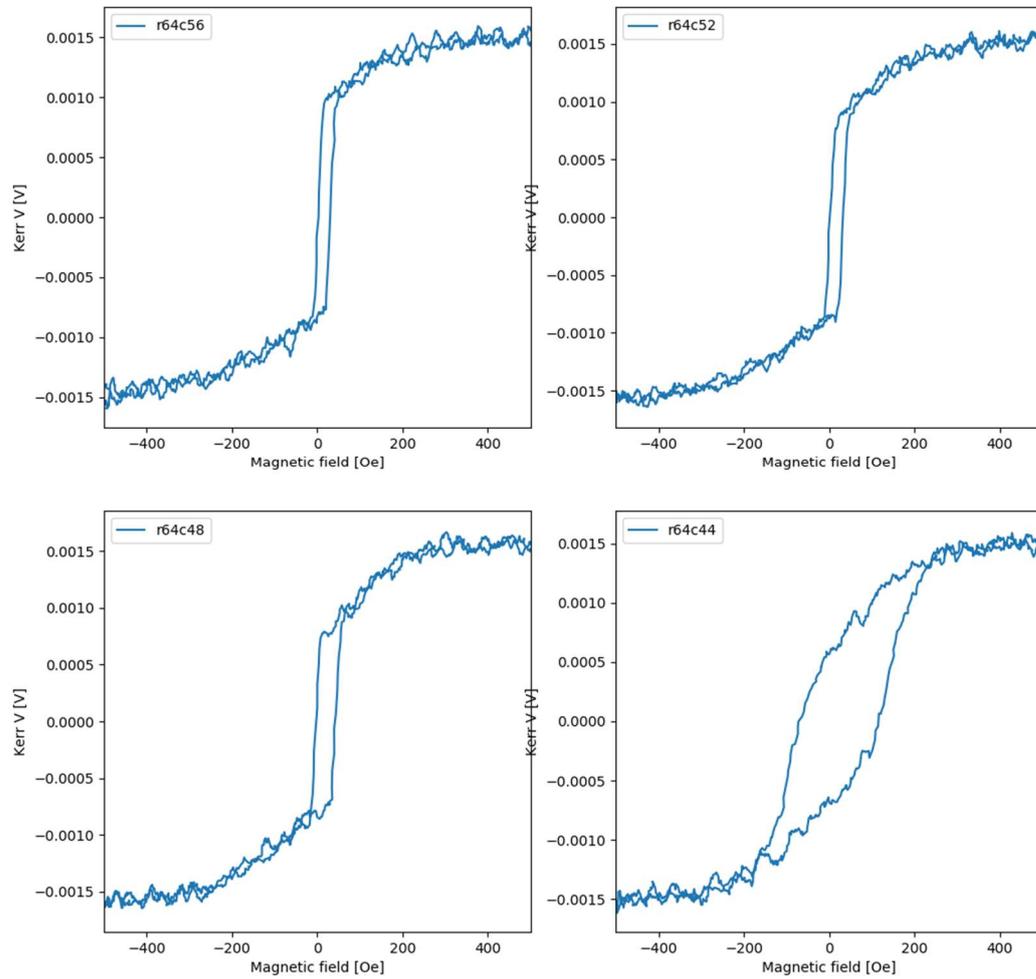


Figure A6: XRD plot for one of the compositions

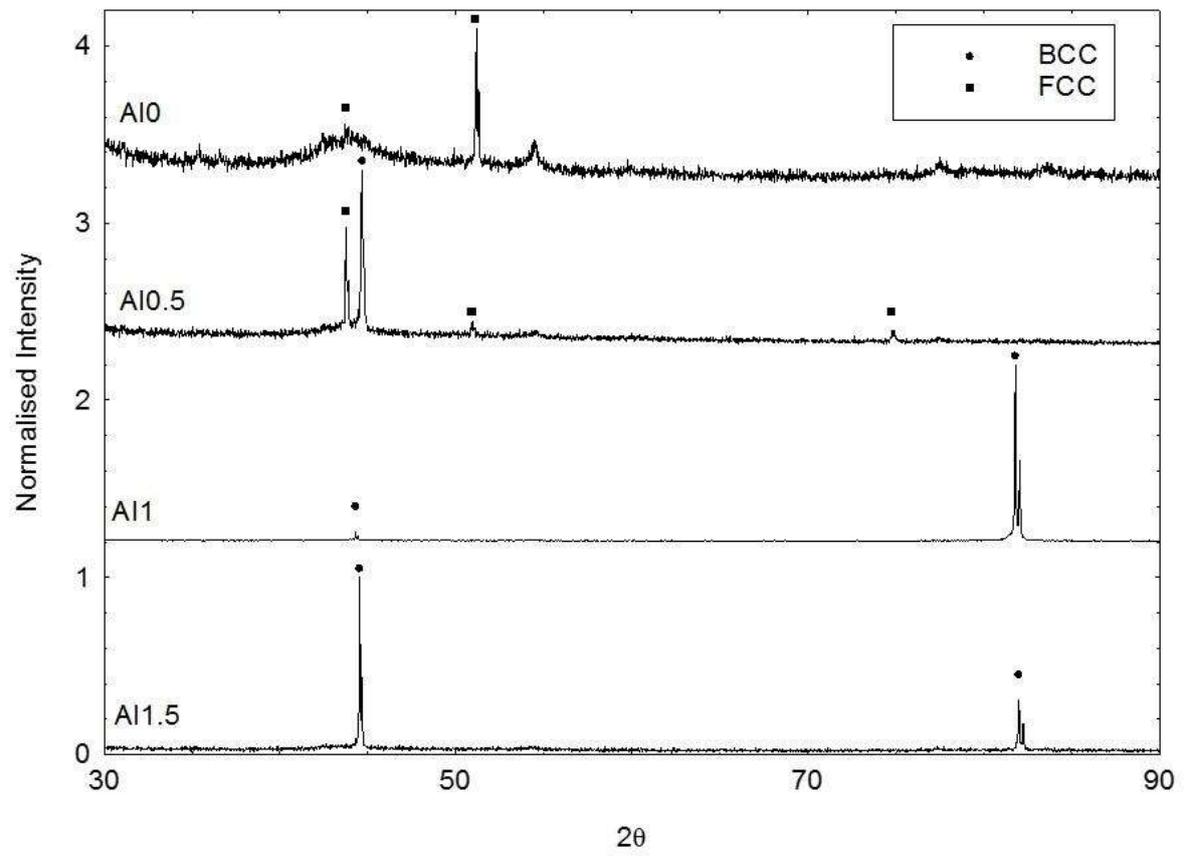
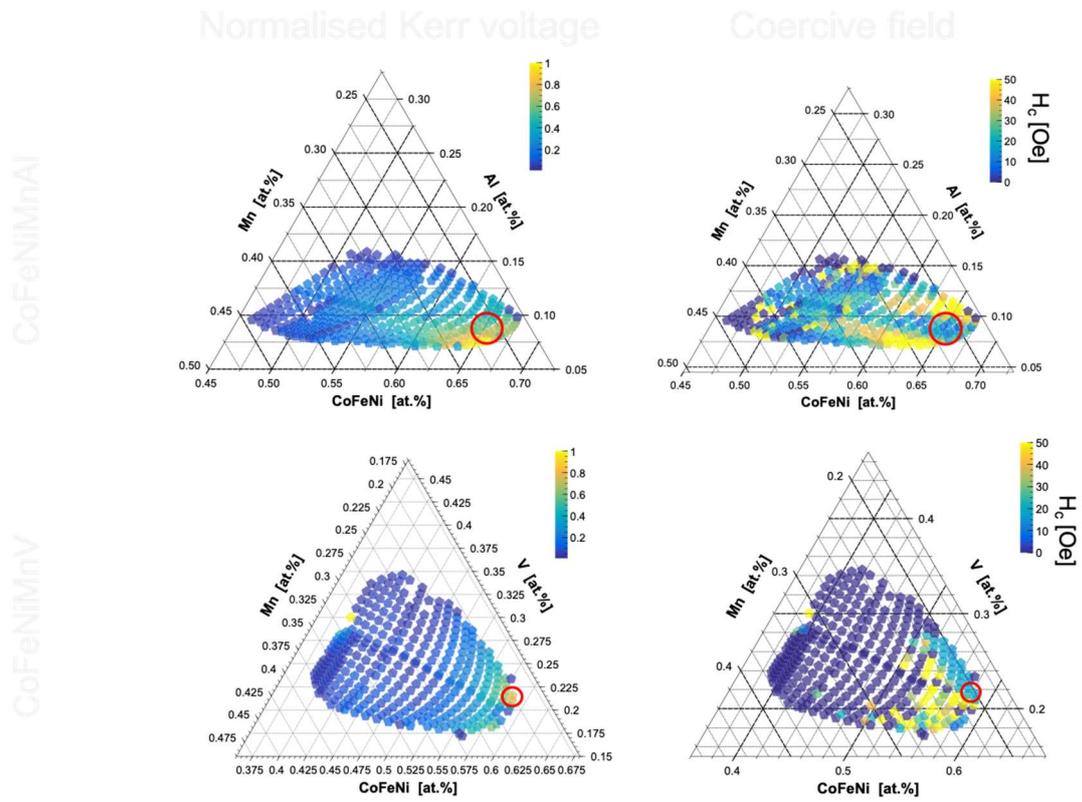


Figure A7: Ternary coercive field and Kerr voltage plots for a single wafer.



## Appendix B

### B1 Example Optical Microscopy “Sample” dataset

Notes:

- Proof-of-concept draft. Not comprehensive.
- Common semantic context with CS4 (or other cases), will be in a shared section.
- JSON-LD context can automatically alter some values to more suited ones.
- Package structure (inner folder tree, if any) respected (as a convenience) but is not the metadata.
- **Is part of** a larger (wafer) dataset with all samples for Optical Microscopy.

```
{
  "@context": {
    "@version": 1.1,
    "@base": "https://psdi-example.com/dataset/",
    "dcat": "http://www.w3.org/ns/dcat#",
    "dc": "http://purl.org/dc/elements/1.1/",
    "dct": "http://purl.org/dc/terms/",
    "dctype": "http://purl.org/dc/dcmitype/",
    "foaf": "http://xmlns.com/foaf/0.1/",
    "locn": "http://www.w3.org/ns/locn#",
    "odrl": "http://www.w3.org/ns/odrl/2/",
    "owl": "http://www.w3.org/2002/07/owl#",
    "prov": "http://www.w3.org/ns/prov#",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "skos": "http://www.w3.org/2004/02/skos/core#",
    "time": "http://www.w3.org/2006/time#",
    "vcard": "http://www.w3.org/2006/vcard/ns#",
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "psdi": "http://www.psd-example.com/psdi#",
    "dqv": "http://www.w3.org/ns/dqv#",
    "Agent": "http://xmlns.com/foaf/0.1/Agent",
    "Checksum": "http://spdx.org/rdf/terms#Checksum",
    "Column": "http://www.w3.org/ns/csvw#Column",
    "Dataset": "http://www.w3.org/ns/dcat#Dataset",
    "Distribution": "http://www.w3.org/ns/dcat#Distribution",
    "Document": "http://xmlns.com/foaf/0.1/Document",
    "Frequency": "http://purl.org/dc/terms/Frequency",
    "Individual": "http://www.w3.org/2006/vcard/ns#Individual",
    "LicenseDocument": "http://purl.org/dc/terms/LicenseDocument",
    "MediaType": "http://purl.org/dc/terms/MediaType",
    "Organization": "http://xmlns.com/foaf/0.1/Organization",
    "PeriodOfTime": "http://purl.org/dc/terms/PeriodOfTime",
    "Person": "http://xmlns.com/foaf/0.1/Person",
    "RightsStatement": "http://purl.org/dc/terms/RightsStatement",
    "Role": "https://www.w3.org/ns/dcat#Role",
    "Schema": "http://www.w3.org/ns/csvw#Schema",
    "Standard": "http://purl.org/dc/terms/Standard",
    "accessRights": {
      "@id": "http://purl.org/dc/terms/accessRights",
      "@type": "xsd:string"
    }
  },
}
```

```
"accessURL": {
  "@id": "http://www.w3.org/ns/dcat#accessURL",
  "@type": "@id",
  "@context": {
    "@base": "https://psdi-example.com/data/",
    "@propagate": false
  }
},
"accrualPeriodicity": {
  "@id": "http://purl.org/dc/terms/accrualPeriodicity",
  "@type": "@id"
},
"algorithm": {
  "@id": "http://spdx.org/rdf/terms#algorithm",
  "@type": "http://spdx.org/rdf/terms#checksumAlgorithm_md5"
},
"byteSize": {
  "@id": "http://www.w3.org/ns/dcat#byteSize",
  "@type": "http://www.w3.org/2001/XMLSchema#decimal"
},
"creator": {
  "@id": "http://purl.org/dc/terms/creator",
  "@type": "@id"
},
"checksum": {
  "@id": "http://spdx.org/rdf/terms#checksum",
  "@type": "@id"
},
"checksumValue": {
  "@id": "http://spdx.org/rdf/terms#checksumValue",
  "@type": "http://www.w3.org/2001/XMLSchema#hexBinary"
},
"conformsTo": {
  "@id": "http://purl.org/dc/terms/conformsTo",
  "@type": "Standard",
  "@context": {
    "label": "http://www.w3.org/2000/01/rdf-schema#label",
    "@propagate": false
  }
},
"contactPoint": {
  "@id": "http://www.w3.org/ns/dcat#contactPoint",
  "@type": "@id"
},
"dataset": {
  "@id": "http://www.w3.org/ns/dcat#dataset",
  "@type": "Dataset"
},
"description": {
  "@id": "http://purl.org/dc/terms/description",
  "@type": "http://www.w3.org/2001/XMLSchema#string"
},
"distribution": {
  "@id": "http://www.w3.org/ns/dcat#distribution",
  "@type": "@id",
  "@context": {
```

```

    "@base": "https://psdi-example.com/data/",
    "@propagate": false
  }
},
"downloadURL": {
  "@id": "http://www.w3.org/ns/dcat#downloadURL",
  "@type": "@id",
  "@context": {
    "@base": "https://psdi-example.com/download/",
    "@propagate": false
  }
},
"hasBeginning": {
  "@id": "https://www.w3.org/TR/owl-time/#time:hasBeginning",
  "@type": "xsd:dateTime"
},
"hasEnd": {
  "@id": "https://www.w3.org/TR/owl-time/#time:hasEnd",
  "@type": "xsd:dateTime"
},
"hadRole": {
  "@id": "https://www.w3.org/ns/dcat#hadRole",
  "@type": "@id"
},
"identifier": {
  "@id": "http://purl.org/dc/terms/identifier",
  "@type": "http://www.w3.org/2001/XMLSchema#string",
  "@container": "@list"
},
"isReferencedBy": {
  "@id": "http://purl.org/dc/terms/isReferencedBy",
  "@type": "@id"
},
"issued": {
  "@id": "http://purl.org/dc/terms/issued",
  "@type": "xsd:dateTime"
},
"keyword": {
  "@id": "http://www.w3.org/ns/dcat#keyword"
},
"landingPage": {
  "@id": "http://www.w3.org/ns/dcat#landingPage",
  "@type": "@id",
  "@context": {
    "@base": "https://psdi-example.com/landingpage/",
    "@propagate": false
  }
},
"language": {
  "@id": "http://purl.org/dc/terms/language",
  "@type": "@id",
  "@context": {
    "@base": "http://id.loc.gov/vocabulary/iso639-1/",
    "@propagate": false
  }
},

```

```
"licence": {
  "@id": "http://purl.org/dc/terms/license",
  "@type": "@id"
},
"mediaType": {
  "@id": "http://www.w3.org/ns/dcat#mediaType",
  "@type": "http://purl.org/dc/terms/MediaType"
},
"modified": {
  "@id": "http://purl.org/dc/terms/modified",
  "@type": "xsd:dateTime"
},
"name": {
  "@id": "http://xmlns.com/foaf/0.1/name",
  "@type": "http://www.w3.org/2001/XMLSchema#string"
},
"publisher": {
  "@id": "http://purl.org/dc/terms/publisher",
  "@type": "@id"
},
"qualifiedAttribution": {
  "@id": "https://www.w3.org/ns/prov#qualifiedAttribution",
  "@type": "@id"
},
"rights": {
  "@id": "http://purl.org/dc/terms/rights",
  "@type": "http://www.w3.org/2001/XMLSchema#string"
},
"temporal": {
  "@id": "http://purl.org/dc/terms/temporal",
  "@type": "PeriodOfTime"
},
"theme": {
  "@id": "http://www.w3.org/ns/dcat#theme",
  "@type": "@id"
},
"title": {
  "@id": "http://purl.org/dc/terms/title",
  "@type": "http://www.w3.org/2001/XMLSchema#string"
},
"tableSchema": {
  "@id": "http://www.w3.org/ns/csvw#tableSchema",
  "@type": "http://www.w3.org/ns/csvw#Schema"
},
"columns": {
  "@id": "http://www.w3.org/ns/csvw#columns",
  "@type": "@id",
  "@container": "@list"
},
"hasQualityMeasurement": "dqv:hasQualityMeasurement",
"QualityMeasurement": "dqv:QualityMeasurement",
"value": "dqv:value",
"isMeasurementOf": "dqv:isMeasurementOf",
"Metric": "dqv:Metric",
"inDimension": "dqv:inDimension",
"precision": "dqv:precision",
```

```

    "expectedDataType": "dqv:expectedDataType",
    "datasetType": "dct:type",
    "hasPart": "dct:hasPart",
    "isPartOf": "dct:isPartOf",
    "wasDerivedFrom": "prov:wasDerivedFrom",
    "wasAttributedTo": "prov:wasAttributedTo",
    "fileTypeString": "xsd:string",
    "subject": "dct:subject",
    "waferRow": {"@id": "psdi:waferRow", "@type": "dqv:QualityMeasurement"},
    "waferColumn": {"@id": "psdi:waferColumn", "@type": "dqv:QualityMeasurement"},
    "coordinateLength": {"@id": "psdi:coordinateLength", "@type": "dqv:Metric"},
    "unitMeasure": "http://purl.org/linked-data/sdmx/2009/attribute#:unitMeasure",
    "id": "@id",
    "type": "@type"
  },
  "id": "8c95605a-ffeb-423c-9267-6c2eea31fd03",
  "type": "Dataset",
  "datasetType": [
    "Sample"
  ],
  "identifier": [
    "8c95605a-ffeb-423c-9267-6c2eea31fd03"
  ],
  "hasQualityMeasurement": [
    {
      "waferRow": {
        "@type": "QualityMeasurement",
        "value": 61,
        "isMeasurementOf": {
          "waferCoordinate": {
            "@type": "Metric",
            "inDimension": "precision",
            "expectedDataType": "xsd:unsignedShort"
          }
        }
      }
    },
    {
      "waferColumn": {
        "@type": "QualityMeasurement",
        "value": 37,
        "isMeasurementOf": {
          "waferCoordinate": {
            "@type": "Metric",
            "inDimension": "precision",
            "expectedDataType": "xsd:unsignedShort"
          }
        }
      }
    }
  ],
  "coordinateLength": {
    "@type": "QualityMeasurement",
    "value": 2,
    "isMeasurementOf": {
      "unitMeasure": "http://qudt.org/vocab/unit/MillIM"
    }
  }
}

```

```

    }
  }
],
"wasAttributedTo": null,
"hasPart": [],
"isPartOf": ["9c95605a-ffeb-423c-9267-6c2eea31fd03"],
"isDerivedFrom": [],
"wasGeneratedBy": null,
"isReferencedBy": null,
"landingPage": "8c95605a-ffeb-423c-9267-6c2eea31fd03",
"language": "en",
"title": "This must be user input. Optical microscope: referring to CoFeNi_50W_Mn_30W_Al_30W_17mins
experiments Region:r61c37",
"description": "User dataset description. Note the folders substructure for the files does not be like this (as
original) but added to the metadata properly and consistently. All coordinates inputs and ouputs could be
aggregated into a dataset \"datasetType\": \"Wafer\" (a dataset series)",
"keyword": [],
"subject": ["output"],
"theme": [],
"conformsTo": [],
"publisher": {
  "id": "https://ror.org/001aqnf71",
  "type": "Organization",
  "name": "UK Research and Innovation"
},
"creator": [
  {
    "id": "https://ror.org/001aqnf71",
    "type": "Organization",
    "name": "UK Research and Innovation"
  }
],
"qualifiedAttribution": [
  {
    "type": "Attribution",
    "hadRole": "funder",
    "agent": {
      "type": "Organization",
      "id": "https://ror.org/0439y7842",
      "name": "Engineering and Physical Sciences Research Council"
    }
  }
],
"contactPoint": {
  "type": "Individual",
  "fn": "Jane Legalitas",
  "hasEmail": "j.legalitas@example.com"
},
"licence": {
  "type": "LicenseDocument",
  "id": "https://spdx.org/licenses/CC-BY-4.0.html",
  "label": "CC-BY-4.0"
},
"rights": "Any literal clarifying comments beyond licence terms.",
"accessRights": "metadata only access",

```

```
"created": "2019-07-09T13:10:00+00:00",
"modified": "2019-07-09T14:10:00+00:00",
"issued": "2019-07-09T14:10:00+00:00",
"temporal": {
  "type": "PeriodOfTime",
  "hasBeginning": "2019-07-09T13:10:00+00:00",
  "hasEnd": "2019-07-09T13:10:00+00:00"
},
"accrualPeriodicity": "http://purl.org/cld/freq/completelyIrregular",
"distribution": [
  {
    "id": "c455f45843eff06437da9fe4d4b02fff",
    "type": "Distribution",
    "title": "N/A",
    "fileName": "./img13.jpg",
    "accessURL": "c455f45843eff06437da9fe4d4b02fff",
    "downloadURL": "c455f45843eff06437da9fe4d4b02fff",
    "mediaType": "image/jpeg",
    "byteSize": "125645",
    "checksum": {
      "type": "Checksum",
      "algorithm": "checksum:Algorithm_md5",
      "checksumValue": "c455f45843eff06437da9fe4d4b02fff"
    },
    "conformsTo": []
  },
  {
    "id": "ea38058d06091de82b06fdb1478869f9",
    "type": "Distribution",
    "title": "N/A",
    "fileName": "./img3.jpg",
    "accessURL": "ea38058d06091de82b06fdb1478869f9",
    "downloadURL": "ea38058d06091de82b06fdb1478869f9",
    "mediaType": "image/jpeg",
    "byteSize": "126153",
    "checksum": {
      "type": "Checksum",
      "algorithm": "checksum:Algorithm_md5",
      "checksumValue": "ea38058d06091de82b06fdb1478869f9"
    },
    "conformsTo": []
  },
  {
    "id": "3892e5a6536b203c363cf31e79922211",
    "type": "Distribution",
    "title": "N/A",
    "fileName": "./img1.jpg",
    "accessURL": "3892e5a6536b203c363cf31e79922211",
    "downloadURL": "3892e5a6536b203c363cf31e79922211",
    "mediaType": "image/jpeg",
    "byteSize": "132602",
    "checksum": {
      "type": "Checksum",
      "algorithm": "checksum:Algorithm_md5",
      "checksumValue": "3892e5a6536b203c363cf31e79922211"
    },
  },
}
```

```
"conformsTo": []
},
{
  "id": "e4250e80630161bcc551c5c786c328a3",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img9.jpg",
  "accessURL": "e4250e80630161bcc551c5c786c328a3",
  "downloadURL": "e4250e80630161bcc551c5c786c328a3",
  "mediaType": "image/jpeg",
  "byteSize": "103400",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "e4250e80630161bcc551c5c786c328a3"
  },
  "conformsTo": []
},
{
  "id": "ee5ec909fe079a373ed5adafa7b501c4",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img12.jpg",
  "accessURL": "ee5ec909fe079a373ed5adafa7b501c4",
  "downloadURL": "ee5ec909fe079a373ed5adafa7b501c4",
  "mediaType": "image/jpeg",
  "byteSize": "113358",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "ee5ec909fe079a373ed5adafa7b501c4"
  },
  "conformsTo": []
},
{
  "id": "c002fb38c6260a01797c4d6469ca040b",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img5.jpg",
  "accessURL": "c002fb38c6260a01797c4d6469ca040b",
  "downloadURL": "c002fb38c6260a01797c4d6469ca040b",
  "mediaType": "image/jpeg",
  "byteSize": "163168",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "c002fb38c6260a01797c4d6469ca040b"
  },
  "conformsTo": []
},
{
  "id": "b272bab7de07dc02504c83c20df08a5e",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img6.jpg",
  "accessURL": "b272bab7de07dc02504c83c20df08a5e",
```

```
"downloadURL": "b272bab7de07dc02504c83c20df08a5e",
"mediaType": "image/jpeg",
"byteSize": "159460",
"checksum": {
  "type": "Checksum",
  "algorithm": "checksum:Algorithm_md5",
  "checksumValue": "b272bab7de07dc02504c83c20df08a5e"
},
"conformsTo": []
},
{
  "id": "5de9d123dc0205bc1cc102a8523c1656",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img7.jpg",
  "accessURL": "5de9d123dc0205bc1cc102a8523c1656",
  "downloadURL": "5de9d123dc0205bc1cc102a8523c1656",
  "mediaType": "image/jpeg",
  "byteSize": "145658",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "5de9d123dc0205bc1cc102a8523c1656"
  },
  "conformsTo": []
},
{
  "id": "ed4ffabc1cc91ba8a2a9a1de208d4588",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img11.jpg",
  "accessURL": "ed4ffabc1cc91ba8a2a9a1de208d4588",
  "downloadURL": "ed4ffabc1cc91ba8a2a9a1de208d4588",
  "mediaType": "image/jpeg",
  "byteSize": "119736",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "ed4ffabc1cc91ba8a2a9a1de208d4588"
  },
  "conformsTo": []
},
{
  "id": "e699e0308bad04b4ddf9005885a36f0e",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img4.jpg",
  "accessURL": "e699e0308bad04b4ddf9005885a36f0e",
  "downloadURL": "e699e0308bad04b4ddf9005885a36f0e",
  "mediaType": "image/jpeg",
  "byteSize": "141646",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "e699e0308bad04b4ddf9005885a36f0e"
  },
}
```

```
"conformsTo": []
},
{
  "id": "36aa4ed8f34bec3d09389c9fa968b7f4",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img10.jpg",
  "accessURL": "36aa4ed8f34bec3d09389c9fa968b7f4",
  "downloadURL": "36aa4ed8f34bec3d09389c9fa968b7f4",
  "mediaType": "image/jpeg",
  "byteSize": "112734",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "36aa4ed8f34bec3d09389c9fa968b7f4"
  },
  "conformsTo": []
},
{
  "id": "adac761bc1f09e31f11d8586b5ba7be6",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img8.jpg",
  "accessURL": "adac761bc1f09e31f11d8586b5ba7be6",
  "downloadURL": "adac761bc1f09e31f11d8586b5ba7be6",
  "mediaType": "image/jpeg",
  "byteSize": "140439",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "adac761bc1f09e31f11d8586b5ba7be6"
  },
  "conformsTo": []
},
{
  "id": "9c46fe262c300499e863c2600b952570",
  "type": "Distribution",
  "title": "N/A",
  "fileName": "./img2.jpg",
  "accessURL": "9c46fe262c300499e863c2600b952570",
  "downloadURL": "9c46fe262c300499e863c2600b952570",
  "mediaType": "image/jpeg",
  "byteSize": "126106",
  "checksum": {
    "type": "Checksum",
    "algorithm": "checksum:Algorithm_md5",
    "checksumValue": "9c46fe262c300499e863c2600b952570"
  },
  "conformsTo": []
}
]
}
```