

Lawrence Berkeley National Laboratory

LBL Publications

Title

A database for storing the results of material radiopurity measurements

Permalink

<https://escholarship.org/uc/item/7mz5m1g6>

Authors

Loach, JC
Cooley, J
Cox, GA
et al.

Publication Date

2016-12-01

DOI

10.1016/j.nima.2016.09.036

Peer reviewed

A Database for Storing the Results of Material Radiopurity Measurements

J.C. Loach^{a,b}, J. Cooley^c, G.A. Cox^d, Z. Li^a, K.D. Nguyen^b, A.W.P. Poon^b

^a*Department of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*

^b*Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

^c*Department of Physics, Southern Methodist University, Dallas, TX 75275, USA*

^d*Institute for Nuclear Physics, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany*

Abstract

Searches for rare nuclear processes, such as neutrinoless double beta-decay and the interactions of WIMP dark matter, are motivating experiments with ever-decreasing levels of radioactive backgrounds. These background reductions are achieved using various techniques, but amongst the most important is minimizing radioactive contamination in the materials from which the experiment is constructed. To this end there have been decades of advances in material sourcing, manufacture and certification, during which researchers have accumulated many thousands of measurements of material radiopurity. Some of these assays are described in publications, others are in databases, but many are still communicated informally. Until this work, there has been no standard format for encoding assay results and no effective, central location for storing them. The aim of this work is to address these long-standing problems by creating a concise and flexible material assay data format and powerful software application to manipulate it. A public installation of this software, available at <http://www.radiopurity.org>, is the largest database of assay results ever compiled and is intended as a long-term repository for the community's data.

1. Introduction

Searches for rare nuclear processes have led to many of the most important results in nuclear and particle physics over recent decades [1, 2, 3, 4] and offer the prospect of many more in years to come [5, 6, 7]. Crucial to these experiments, and especially to those that search for rare weak interactions, such as neutrinoless double beta-decay and the interactions of WIMP dark matter, is the suppression of events due to radioactive backgrounds. This can be achieved through experimental design or analysis, but the most basic strategy is to directly address the source terms by shielding the experiment from environmental radioactivity and, of particular relevance here, constructing the experiment from low-radioactivity materials.

Selecting candidate low-radioactivity materials is as much an art as it is a science, relying on inference from previous measurements and on the instinct of the experimentalist. It is also a critical task because the testing and certification of candidate materials, though a well-established process, involves precision measurements that consume significant amounts of time, cost and effort. Candidates must be selected judiciously and how well this can be done depends upon the quality of information available to the researcher. Of particular importance is their access to previous measurements of similar materials.

A central repository of material radiopurity measurements is strongly motivated by these considerations, and the authors are not the first to propose one or, even, to build one. Previous efforts such as the public material assay database created by the ILIAS collaboration [8] have been invaluable to the low-background physics community. But despite their usefulness, none have fully met the community's needs. There have been issues with the limited scope of the data sets and difficulties in

augmenting them, as well as with non-portability of data and difficulties in querying it. Solving these issues was the motivation for this work, and we describe ways of storing and handling material assays that are intended to address them.

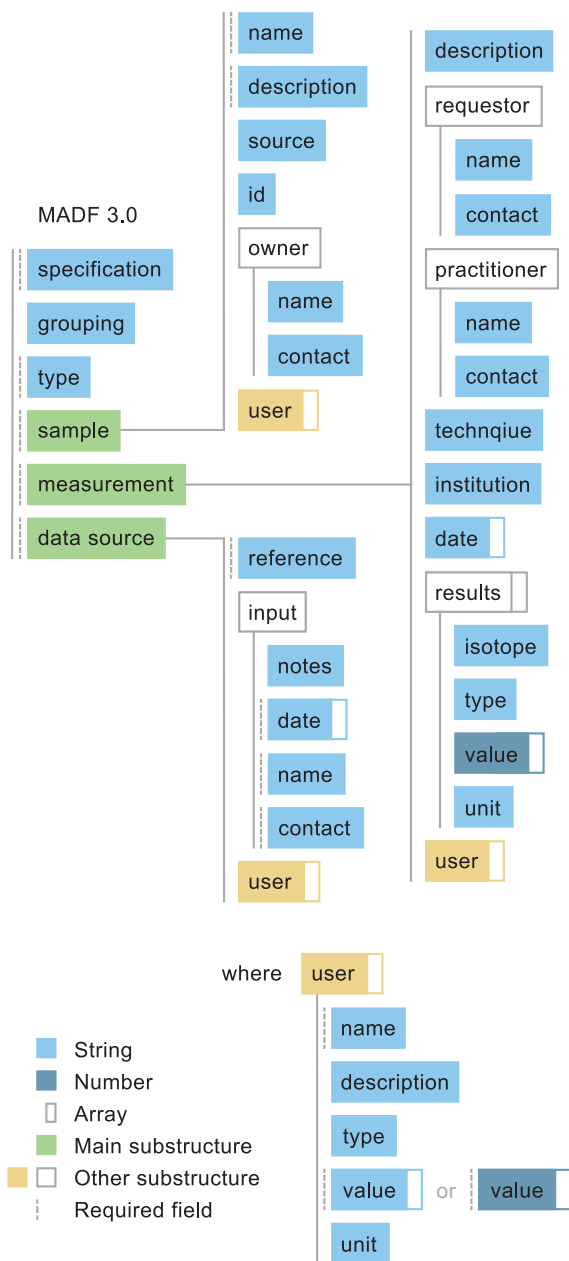
Our work comprises two parts: a data format in which assays can be encoded and a piece of software for manipulating the encoded data. In a practical sense these components are intertwined, but there is a formal separation between them that is reflected in the organization of this paper. Our material assay data format is presented in the next section, free of association with any particular database software. Following this we discuss our web application Persephone [9], which is a tool for storing, viewing and manipulating data encoded in the format. And, finally, we describe a specific installation of our software that we intend as the central repository for the community's data.

2. The Material Assay Data Format

The Material Assay Data Format (MADF) describes a way to encode assays of material radiopurity in JavaScript Object Notation (JSON) [10]. JSON is a language-independent, open standard for encoding structured data. It is especially widespread in Internet communications and is natively read by many advanced and popular computational tools, including those used in physics analyses.

Each MADF JSON document represents a single assay performed on a particular sample of material. It contains three main substructures: `sample` describes the sample of material being assayed; `measurement` describes the assay process and

Table 1: The structure of the MADF 3.0 data format. All fields are strings unless stated otherwise. Fields marked with a * are required.



Field	Description
specification*	MADF specification version.
grouping	Experiment name or similar.
type*	Fixed value "assay" indicating the document type.
sample	
.name*	Concise description.
.description*	Detailed description.
.id	Identification number.
.source	Where the sample came from.
.owner	Who owns the sample.
.name	Name.
.contact	Email address or telephone no.
.user	User field array. See below.
measurement	
.description	Detailed description.
.requestor	Who coordinated the measurement.
.name	Name.
.contact	Email address or telephone no.
.practitioner	Who did the measurement.
.name	Name.
.contact	Email address or telephone no.
.technique	Technique name.
.institution	Institution name.
.date	Array of strings. See Appendix C.
.results	Array of measurements.
.isotope	Isotope name, usually in the format symbol-mass number.
.type	measurement, limit or range.
.value	Array of one to three numbers. See Appendix A.
.unit	Unit chosen from a restricted set of choices. See Appendix B.
.user	User field array. See below.
data_source	
.reference*	Where the data came from.
.input	Data entry details.
.notes	Input simplifications, assumptions.
.date*	Array of strings. See Appendix C.
.name*	Name.
.contact*	Email address or telephone no.
.user	User field array. See below.
.user	
.name*	Concise description, usually a single word.
.description	Detailed description.
.type	measurement, limit, range or string.
.value*	String or an array of one to three numbers. See Appendix A.
.unit	Unit. See Appendix B.

Figure 1: The structure of the MADF 3.0 data format. Each entry in a user array has the structure given in the lower-right part of the figure. These arrays are used to add extra, use-case-specific fields to the core data structure. The data format is expressed in JSON. Further details are given in Table 1.

its results, including a list of isotopes with their measured values or limits; and `data_source` gives the origin of the information encoded in the document and describes how it was encoded. Overviews of MADF 3.0 are shown in Figure 1 and Table 1. Further details are given in the appendices.

The design principles behind the data format are summarized as follows:

Simplicity. Each assay is encoded in a single, semi-structured, human-readable JSON document and there are no structural relationships between documents. In contrast to relational databases, with their arrangements of tables, all information is encapsulated in a single piece of data. This makes the data highly portable, and assays can even be output and stored in self-contained text files.

In MADF many fields can store extended, descriptive information and only a small number of fields are mandatory. In this way it is more similar to a structured experimental report than, for example, a line in a spreadsheet. This simplicity helps the process of encoding data because there is no need to parse data from disparate sources into a highly-structured form. It helps makes the process amenable to automation and minimizes the decisions required during data entry (and therefore the required skill level of the people doing the entry). One consequence of a simple data format is that it tends to make querying data more difficult but this problem is eased by modern tools such as MapReduce [11] which are implemented in databases designed to store JSON documents.

Inclusivity. MADF tries to minimize constraints on what can be entered in particular fields. For example, there are no requirements that measured values have uncertainties or that limits have confidence levels. The rationale is that more information is generally better, and that the intended users will be sufficiently qualified to judge what is relevant for their application. A measured value without an explicit uncertainty might not be appropriate in a scientific publication but it is not without meaning in certain contexts. However, expecting the user to make judgements only works when the level of information provided is sufficient and, in particular, when they can understand the context and origin of the data. MADF therefore contains required fields for storing a reference to the source data and information on data entry, including the contact details of the person who entered the data. There is also an optional field for describing any assumptions or simplifications made during data entry.

Flexibility. The core of MADF is a simple, compact structure, but it can be extended arbitrarily to accommodate specific needs. A typical use case might involve extra fields for information that is important to people who perform assays, but less so to those who use the results. Examples might be file names, calibration data, images, energy spectra or even analysis code (stored as serialized objects or, in the case of Python, as IPython Notebook JSON objects). The data

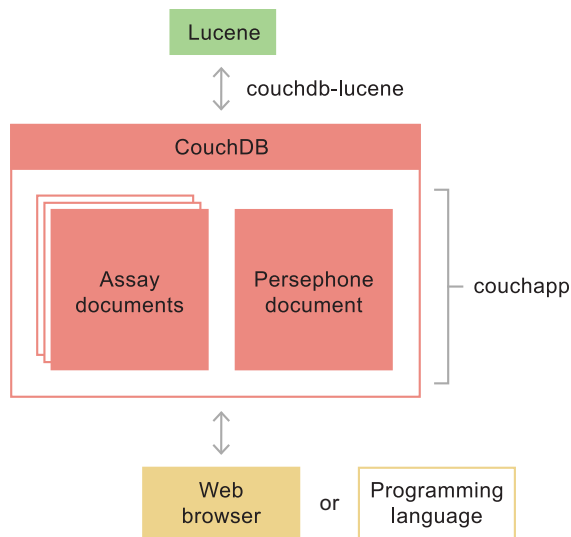


Figure 2: The structure of the database and web application. Gray arrows indicate HTTP requests.

structure can be extended in two ways: freely or through user arrays. Free extension means adding arbitrary fields anywhere within the structure. In contrast, user arrays contain predefined structures that can be added only to the three main `sample`, `measurement` and `data_source` substructures. The difference between them is that all applications implementing MADF should recognize and display user arrays, while there are no requirements on how they should handle free extensions. Our application Persephone, discussed below, ignores free extensions.

The structure of MADF is described by a JSON schema [12], against which documents can be validated using open source libraries (such as [13] for Python). Future development of MADF is encouraged by the independence of the schema from the data representation, which means that changes can be made with substantial freedom and with few consequences for existing data. One reason for the failure of previous databases has been their structural rigidity, which has made it difficult to add new data and difficult to adapt to new forms of data. A database using JSON documents is maximally flexible and adaptable to future needs.

JSON documents can be stored in a variety of database systems. These include databases specially designed around the format, as well as traditional relational databases. We discuss one database system in the following section in the context of our web application.

3. Persephone

Persephone [9] is an open-source software package for storing, displaying and manipulating assays encoded in MADF. It is a JavaScript-based web application stored within, and served by, an Apache CouchDB database [14]. The architecture is illustrated in Figure 2.

CouchDB is an open-source, document-oriented database which uses JSON documents for data storage. It is schema-less, meaning that it does not enforce any pre-defined structure on its documents or on the relationships between them; schemas are specified and enforced by user-developed code that interacts with the database. CouchDB uses MapReduce as its query language and HTTP as its primary API. It runs as a web server and provides sophisticated functionality for replicating databases between different locations and guaranteeing eventual consistency. CouchDB does not currently have native search capabilities but it can interface with other software such as Apache Lucene [15, 16] to provide this functionality.

These features make CouchDB an ideal tool for a material assay database (and, indeed, for a variety of applications in physics experiments [17, 18]). It provides the flexibility required by the data format and its use of HTTP makes it easy to interface with using most programming languages. In addition, being both schema-less and a web server, CouchDB has the ability to store entire client-side web applications as documents within the database and serve them to users. Tools such as couchapp [19] are available to help encode web applications as JSON documents. Persephone is an application of this type.

Persephone is a client-side web application stored inside a CouchDB database and served by it. It is built from HTML, CSS and JavaScript using standard web programming techniques, such as AJAX, and standard libraries, such as jQuery [20], jQueryUI [21] and the CouchDB jquery plugin. The user interface is a single web page containing tabs for search, data entry, data editing, settings and user login.

The search interface allows standard search grammar such as wild cards, logical operators and named fields. Search results are initially shown in an abbreviated form, where they can be sorted or removed from the display. The user can expand the abbreviated forms to two deeper levels of detail. Each entry has a tools menu that allows the user to report errors, export the entry (in JSON, CSV, XML or HTML) or, if they have appropriate permissions, invoke the data editing interface.

The data entry interface is a HTML form that allows assays to be entered in a user-friendly way. It supports the whole MADF specification with the exception of free extensions (the addition of fields outside the core data structure and user arrays). Once submitted, data is validated against the JSON schema and then stored in the database. The interface for data editing is similar in appearance and function.

The settings tab allows the user to set persistent values for the email address used to report errors and for the number of search results that are displayed by default.

The login tab provides an interface to CouchDB's user authentication features, which allow simple management of usernames, passwords and read/write permissions.

Persephone is accompanied by a set of python tools that provide command line functionality for submitting, downloading, editing and validating documents. Python provides excellent support for JSON using the json module. The jsonschema module is used to handle validation and the couchdb module is used to communicate with the database. These tools allow data to be downloaded into text files and uploaded from them, and pro-

vide an alternative way to edit existing data and submit new data. The scripts can be incorporated into larger codes to assist automatic conversion of data from other sources.

The application is designed to be run in a number of different contexts: by private individuals, by institutions who perform material assays, by experimental collaborations, and also as a central public database of community data. It can be installed on local instances of CouchDB, on public-facing servers, or in the cloud using commercial services, with databases easily replicated between installations. It could be adapted to run on mobile devices. This model allows an experiment or institution to easily share their internal data with the wider community (by replicating the contents of their local database to the public database). The distributed nature of this model should also help protect the community's data by preventing reliance on one individual or institution as a custodian.

4. The Community Database

A public instance of Persephone exists at <http://www.radiopurity.org>. This database is hosted and managed by SNOLAB and is intended to be a long-term repository for the community's assay data. It currently holds more than 1000 assays, including measurements from peer-reviewed publications [22, 23, 24, 25] and the ILIAS database. Published data are added as the authors become aware of them (there is a contact form on the website) and ways to incorporate more historical data are being explored. Encoding historical data can be a substantial undertaking and the amount of this data that ends up in the database will be strongly dependent on the availability of resources. We encourage the community to support this valuable activity.

5. Conclusion

MADF, Persephone and the community database are three components of a long-term solution to the problem of effectively storing, sharing and searching the results of material assays. MADF provides a flexible, human-readable, non-proprietary data format for encoding material assays as JSON documents. These documents can be stored in a CouchDB database or easily converted to other formats for use elsewhere. Persephone is an open-source web application for manipulating assays encoded in MADF. It is stored inside a CouchDB database, which also serves it to the user. Persephone allows data to be searched, submitted and edited in a powerful and convenient way. The community installation of Persephone at <http://www.radiopurity.org> is intended as the long-term repository for the community's assay data.

6. Acknowledgements

This work was supported by: the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under Contract No. DE-AC02-05CH11231; by the Shanghai Key Lab for Particle Physics and Cosmology (SKLPPC), Grant No.

11DZ2260700; by the Helmholtz Alliance for Astroparticle Physics (HAP) funded by the Initiative and Networking Fund of the Helmholtz Association; and by the National Science Foundation under Grant No. PHY 124260 awarded to the AARM (Assay and Acquisition of Radiopure Materials) collaboration. We acknowledge the support of IBM Cloudant in hosting an early version of our community database and to SNOLAB for their ongoing hosting and management of the site. We acknowledge support and guidance from colleagues in the underground science community, including from the AARM, ILIAS, EXO, XENON100, SuperCDMS and MAJORANA collaborations.

Appendix A. MADF value arrays

value arrays appear in the `measurement.results` sub-structure and in the `user` arrays. In both of these, the array may contain one to three numbers, whose meanings can be inferred from the measurement type, as described below. In addition, in the `user` arrays, the value array may also contain a single string.

type, value	Description
measurement	
[0]	Central value, no error.
[0,0]	Measurement, symmetric error.
[0,0,0]	Measurement, asymmetric error (+, -).
range	
[0,0]	Range (lower, upper).
[0,0,0]	Range (lower, upper), confidence level.
limit	
[0]	Upper limit.
[0,0]	Upper limit with confidence level.
string	
[""]	Meaning undefined.

Below, by way of illustration, is an example `results` array encoding a positive measurement of U-238 and a limit on Th-232:

```
"results": [
  {
    "isotope": "U-238",
    "type": "measurement",
    "value": [400, 20],
    "unit": "ppb"
  },
  {
    "isotope": "Th-232",
    "type": "limit",
    "value": [100, 90],
    "unit": "ppt"
  }
]
```

Appendix B. MADF units

unit strings appear in the `measurement.results` sub-structure and in the `user` arrays. In the `user` arrays there is no restriction on their contents but in the `measurement.results` sub-structure they must contain a value from this list:

- pct (percent by mass)
- g/g, ppm, ppb, ppt, ppq (parts per quadrillion)
- g, mg, ug (micrograms), ng, pg
- Bq, mBq, uBq, nBq, pBq
- X/Y where X is any of g, mg, ug, ng, pg, Bq, mBq, uBq, nBq, pBq and Y is any of unit, kg, cm, m, cm2, m2, cm3, m3 (where 2 and 3 represent powers)

Appendix C. MADF dates

date arrays appear in the `measurement` and `data_source.input` sub-structures. The array can be empty or else contain one or two strings in the format YYYY-MM-DD, YYYY-MM or YYYY. A single string represents a date and two string represent a range of dates. Examples are:

```
date = []
date = ["1980-04"]
date = ["1980-04-26", "1980-04-28"]
```

References

References

- [1] Q. R. Ahmad et al., Direct evidence for neutrino flavor transformation from neutral-current oscillation. *Phys. Rev. Lett.* 89 (2002) 011301. doi:10.1103/PhysRevLett.89.011301. URL <http://link.aps.org/doi/10.1103/PhysRevLett.89.011301>
- [2] T. Araki et al., Measurement of neutrino oscillation with KamLAND: Evidence of suppression. *Phys. Rev. Lett.* 94 (2005) 081801. doi:10.1103/PhysRevLett.94.081801. URL <http://link.aps.org/doi/10.1103/PhysRevLett.94.081801>
- [3] M. Agostini et al., Results on neutrinoless double- β decay of ^{76}Ge from phase i of the. *Phys. Rev. Lett.* 111 (2013) 122503. doi:10.1103/PhysRevLett.111.122503. URL <http://link.aps.org/doi/10.1103/PhysRevLett.111.122503>
- [4] R. Agnese et al. (SuperCDMS Collaboration), WIMP-search results from the second CDMSlite run, Submitted to *Phys. Rev. Lett.* arXiv:1509.02448.
- [5] D. Akerib, et al., The Large Underground Xenon (LUX) experiment, *Nucl. Instrum. Meth.* A704 (2013) 111–126.
- [6] N. Abgrall et al. (MAJORANA Collaboration), The MAJORANA DEMONSTRATOR neutrinoless double-beta decay experiment, *AHEP* (2014) 365432.
- [7] D. R. Artusa et al., Searching for neutrinoless double-beta decay of ^{130}Te with CUORE, *Advances in High Energy Physics* doi:doi:10.1155/2015/879871.
- [8] ILIAS database on radiopurity of materials, <http://radiopurity.in2p3.fr> (accessed 2016-04-13).
- [9] Persephone, <http://github.com/radiopurity/persephone> (accessed 2016-04-13).
- [10] C. Severance, Discovering JavaScript Object Notation, *Computer* 45 (4) (2012) 6–8. doi:10.1109/MC.2012.132.

- [11] J. Dean, S. Ghemawat, MapReduce: Simplified data processing on large clusters., Proc. of the 6th OSDI (2004) 137–150.
- [12] JSON schema, <http://json-schema.org> (accessed 2016-04-13).
- [13] jsonschema, <http://pypi.python.org/pypi/jsonschema> (accessed 2016-04-13).
- [14] Apache CouchDB, <http://couchdb.apache.org> (accessed 2016-04-13).
- [15] Apache Lucene, <http://lucene.apache.org> (accessed 2016-04-13).
- [16] couchdb-lucene, <http://github.com/rnewson/couchdb-lucene> (accessed 2016-04-13).
- [17] G.A. Cox et al., A multi-tiered data structure and process management system based on ROOT and CouchDB, Nucl. Instrum. Meth. A 684 (2012) 63 – 72. doi:<http://dx.doi.org/10.1016/j.nima.2012.04.049>.
- [18] N. Abgrall et al. (MAJORANA Collaboration), The MAJORANA parts tracking database, Nucl. Instrum. Meth. A 779 (2015) 52 – 62. doi:<http://dx.doi.org/10.1016/j.nima.2015.01.001>.
- [19] couchapp, <http://github.com/couchapp/couchapp> (accessed 2016-04-13).
- [20] jQuery, <https://www.jquery.com> (accessed 2016-04-13).
- [21] jQueryUI, <https://www.jqueryui.com> (accessed 2016-04-13).
- [22] D.S. Leonard et al., Systematic study of trace radioactive impurities in candidate construction materials for EXO-200, Nucl. Instrum. Meth. A 591 (3) (2008) 490 – 509. doi:[10.1016/j.nima.2008.03.001](http://dx.doi.org/10.1016/j.nima.2008.03.001).
- [23] E. Aprile et al., The XENON100 dark matter experiment, Astropart. Phys. 35 (9) (2012) 573 – 590. doi:[10.1016/j.astropartphys.2012.01.003](http://dx.doi.org/10.1016/j.astropartphys.2012.01.003).
- [24] G. Alimonti et al., Ultra-low background measurements in a large volume underground detector, Astropart. Phys. 8 (3) (1998) 141 – 157. doi:[10.1016/S0927-6505\(97\)00050-9](http://dx.doi.org/10.1016/S0927-6505(97)00050-9).
- [25] N. Abgrall et al. (MAJORANA Collaboration), The MAJORANA DEMONSTRATOR radioassay program, Submitted to Nucl. Instrum. Meth. AarXiv:1601.03779.