NOMAD Laboratory: Metadata

Fawzi Mohamed

26.1.2016

NOMAD
NOVEL MATERIALS DISCOVERY

What is meta data

- data on the data
- here data is values from inputs and outputs of atomistic calculations
- the key if data is stored in a dictionary
- dictionary: clearly define what is meant with a given name

Meta Data: Aims

- describe *all* data in the inputs and outputs of atomistic simulation programs
- enable analysis of data possibly produced by others
- enable querying of the collected data
- conceptual model for our data
  - define how the data that we extract is organized, and what it is
  - important both for human and for the machine
- make parsing simpler

NOMAD
NOVEL MATERIALS DISCOVERY

Nomad Meta Data...(1/2)

- metadata becomes more useful the more it is understood and used
  - machine readable description
  - tools to visualize it for humans
  - open to external contributions
- code independent part, with code dependent extensions
- not just dictionary, but also some relationships: storage and queries

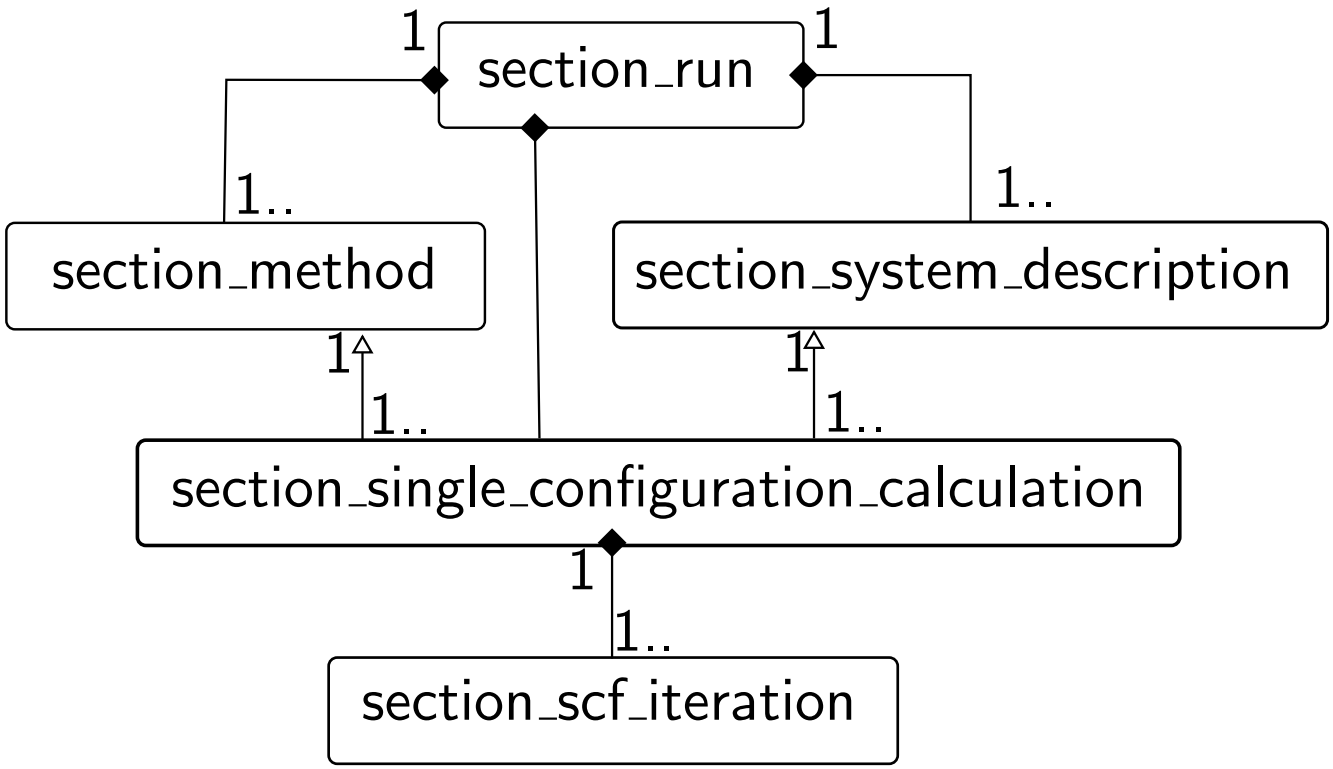
NOMAD
NOVEL MATERIALS DISCOVERY

Nomad Meta Data...(2/2)

- data values consist of basic data types and multidimensional arrays of them
- group together similar *types* making them inherit from the same abstract type (all energies inherit from *energy_value*)
- group together *values* with sections
  - ... but try to avoid unnecessary grouping (try to be flat)
- allow references between sections
- description at `https://gitlab.mpcdf.mpg.de/nomad-lab/public-wiki/wikis/nomad-meta-info`

NOMAD
NOVEL MATERIALS DISCOVERY

Common meta data: core sections



Comparing Calculations

- browse `https://nomad-dev.rz-berlin.mpg.de/ui/index.html`
- method
- basis_set
- XC_method
- transformation: all values in section method that are settings_XC

Queries we want to make possible

- collect program_name XC_method, energy_total, atom_position, atom_label
  of the lowest energy_total for each XC_method and atom_kind and program_name
  in periodic system involving only two atom types.
- combine this with queries using structural similarity (lowest 100 values that are at least X "distance" from each other)

Standard formats

- meta info also defines standard formats
- json (section = dictionary), first level subsections in subsections array
- netcdf (hdf5)