Materials Project Website Walkthrough

Matt Horton
Materials Project Staff, Lawrence Berkeley National Lab

The Materials Project
materialsproject.org

Purpose of this session

SLIDES

Introduction so we're all on the same page

Understand the data that the Materials Project offers and how this data is organized.

DEMO

Learn how to navigate the Materials Project website

In later sessions we will re-visit some of the data and tools shown on the website but using Python within the CoCalc teaching environment.

Where does Materials Project data come from?

The Materials Project is a database of predicted properties of materials using Density Functional Theory (DFT). Teaching DFT is not a goal of this workshop, one of the reasons MP exists is so that you don't need to run DFT yourself!

However, there are some important things you should know DFT:

1. **DFT is an atomistic method** → we need approximate atomic positions and lattice parameters.

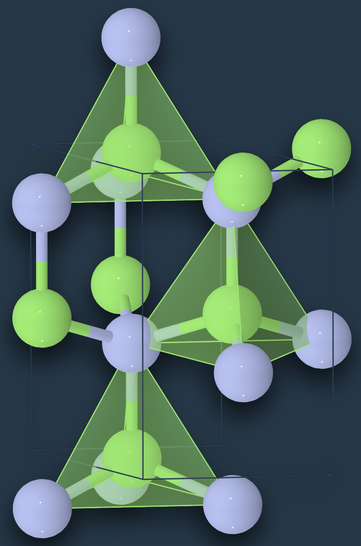
2. **DFT is a ground-state theory** → calculations are at zero temperature and pressure, excited state properties have larger error (especially band gap).

3. **DFT is *mostly* first-principles** → predictions equally good for new and existing materials.

4. **DFT is not entirely first-principles** → most errors are *systematic* in nature, empirical correction can be used to reduce these errors (e.g. in formation enthalpy)

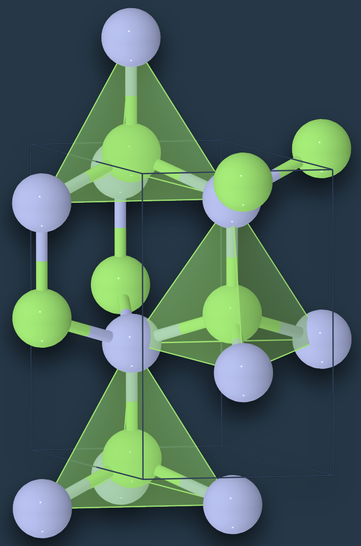
Our documentation and references therein contain more detailed information for specific properties.

How does the Materials Project organizes calculations?



Crystal structure arrives via third-party database, user submission, algorithmic generation, etc...

How does the Materials Project organize its calculations?



Crystal structure arrives via third-party database, user submission, algorithmic generation, etc...

We then perform a sequence of calculations for that crystal structure

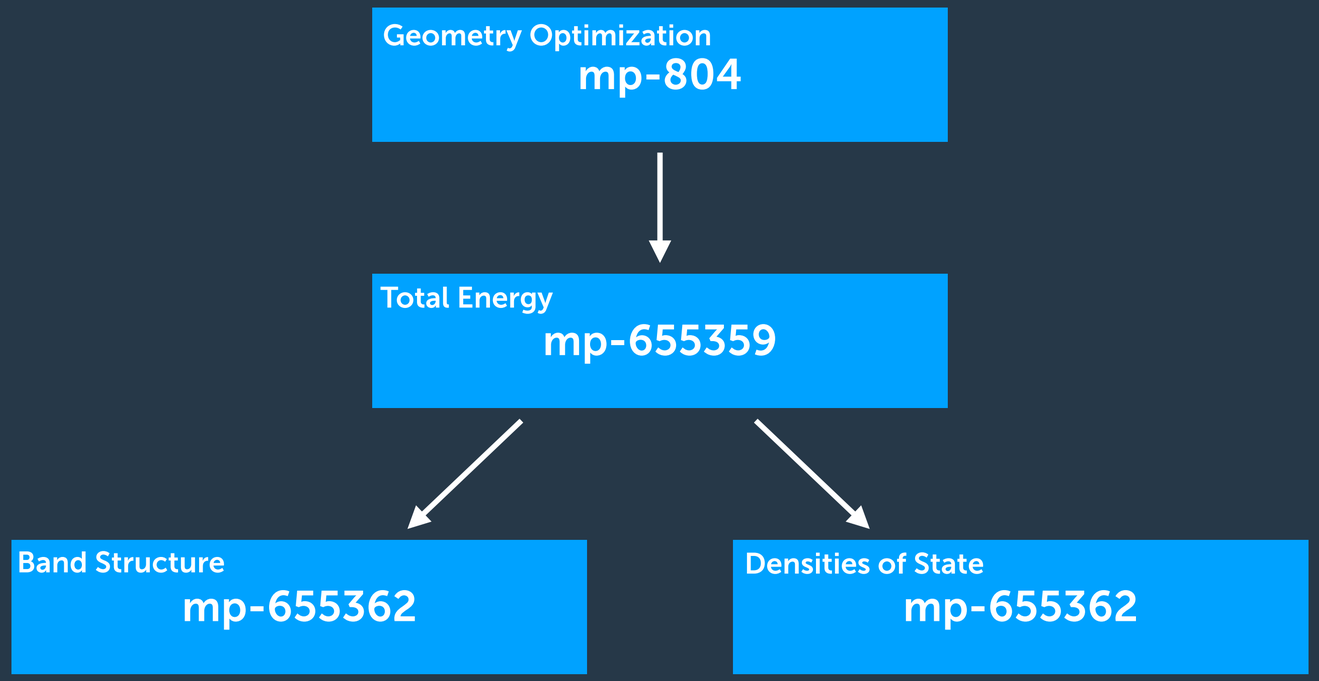
Geometry Optimization

Total Energy

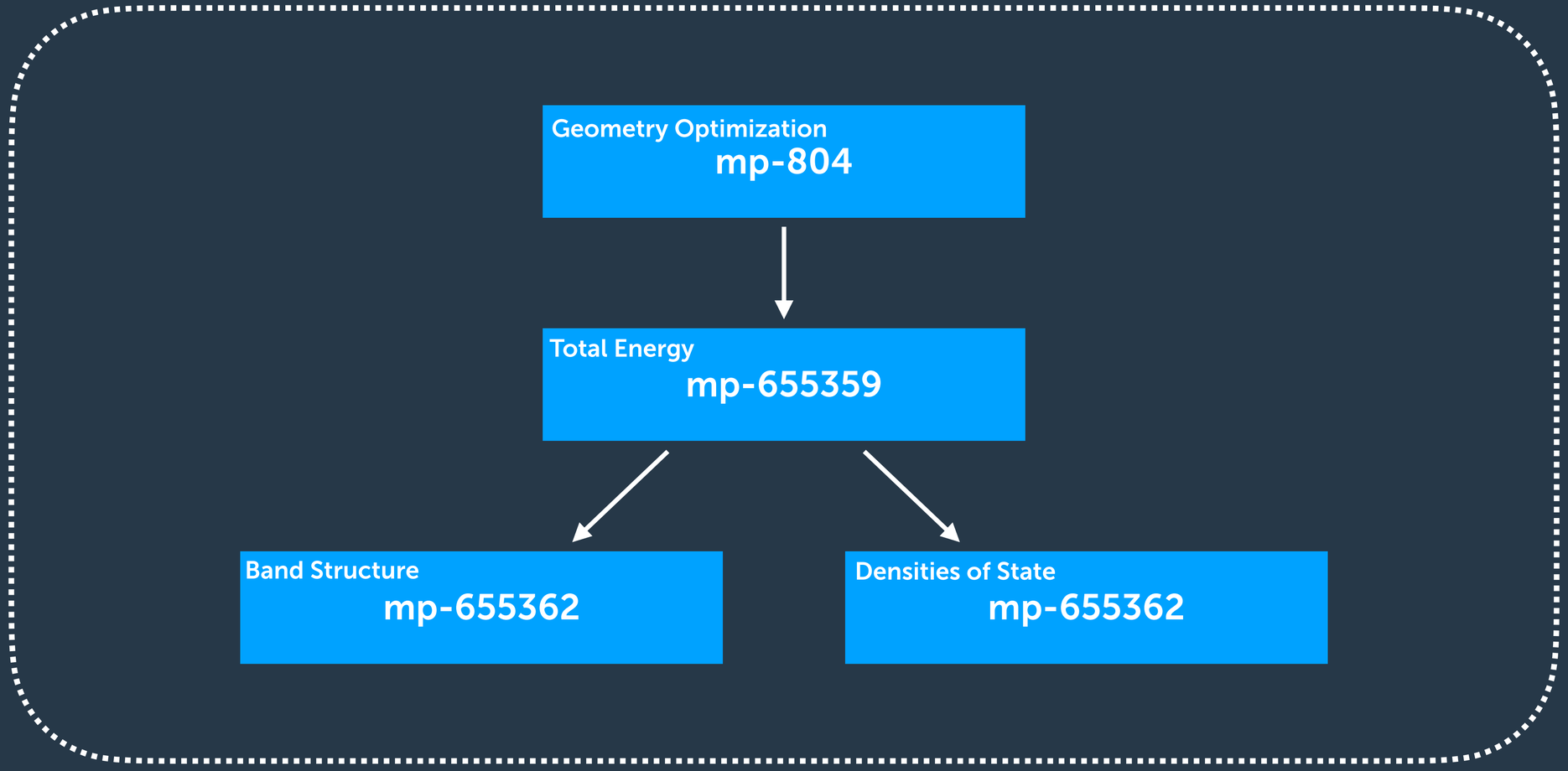
Band Structure

Densities of State

To keep track of calculations a Materials Project ID, or mp-id, gets assigned for every calculation *task*

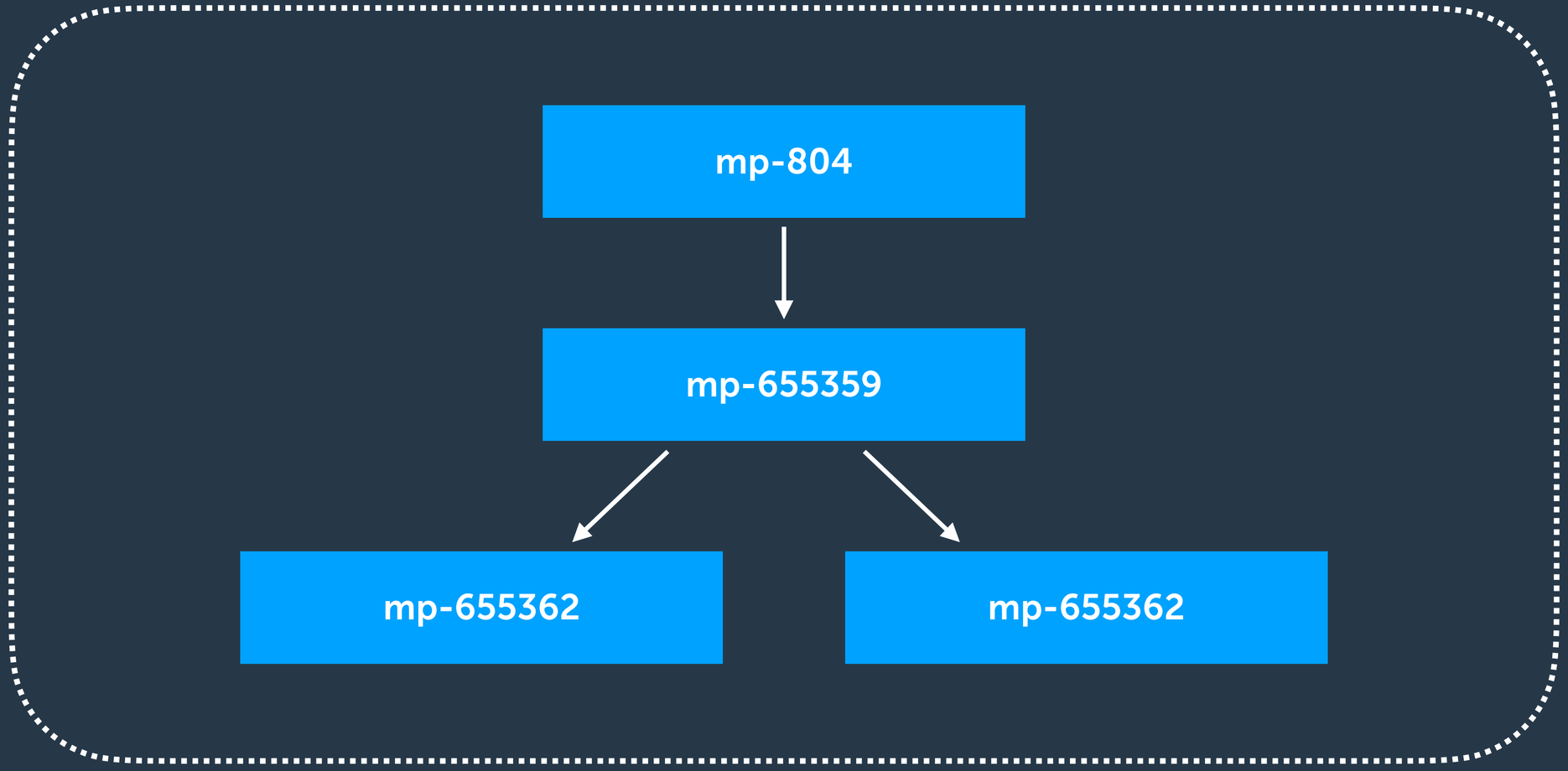


“A material” is simply then a set of these calculation tasks grouped together



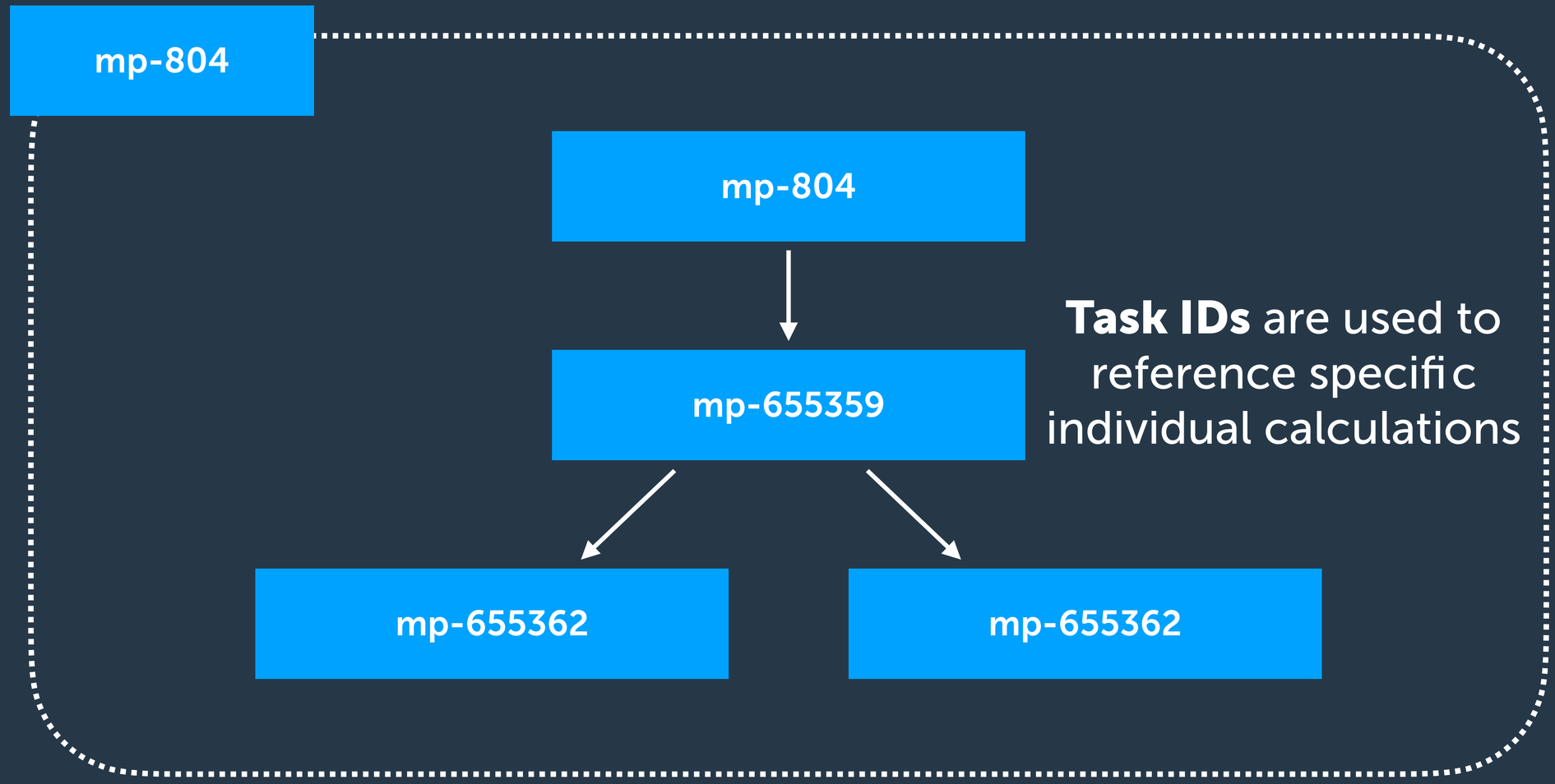
Grouping of calculations is performed using “StructureMatcher,” a feature of *pymatgen* that identifies approximate equivalence between crystal structures

To identify the material, we elevate the smallest mp-id to be its canonical identifier and call this the “material ID”



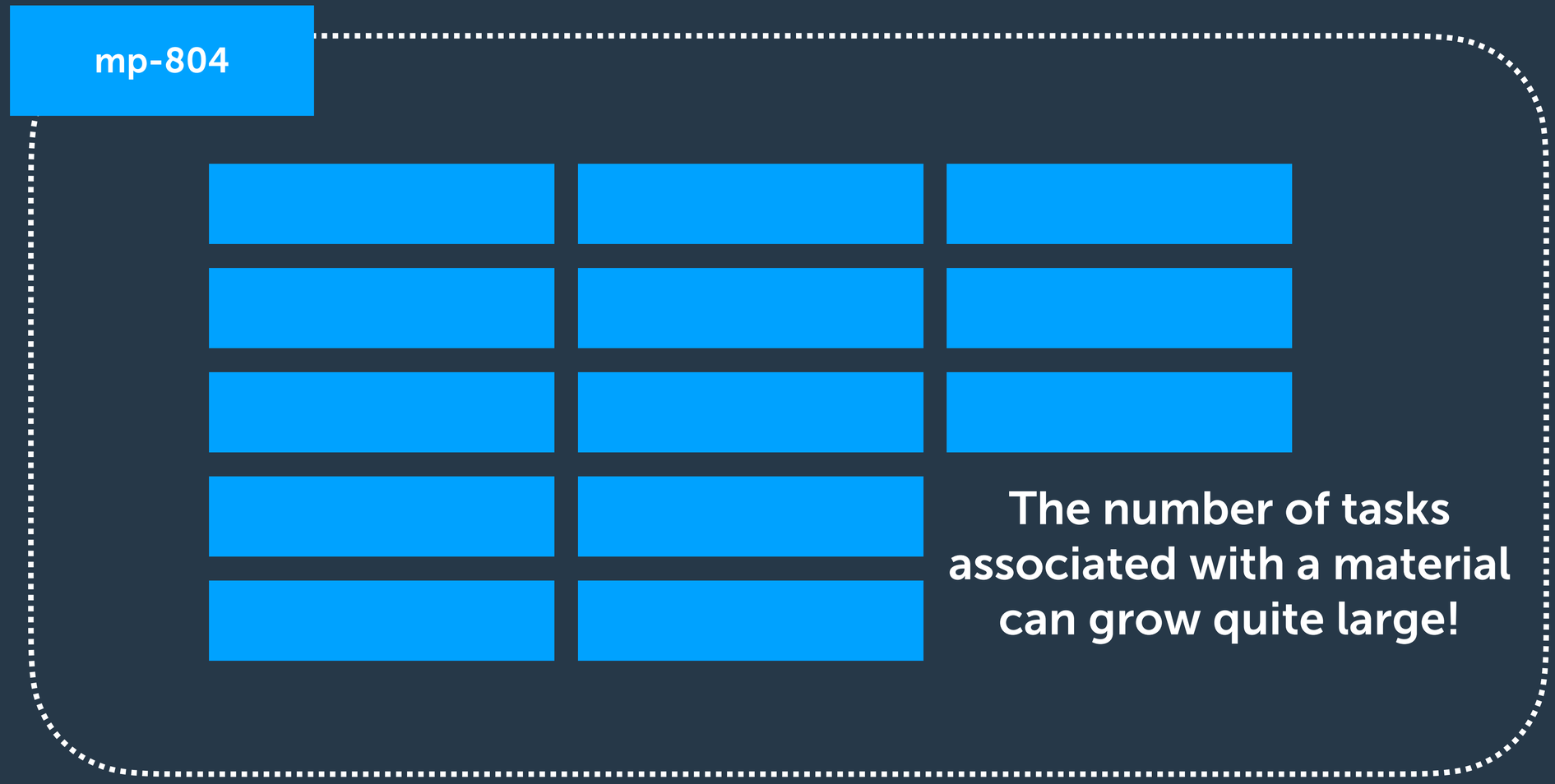
Most users of data from Materials Project will only use or see the “material ID.”

The **Material ID** is used for citing a material and in Materials Project URLs

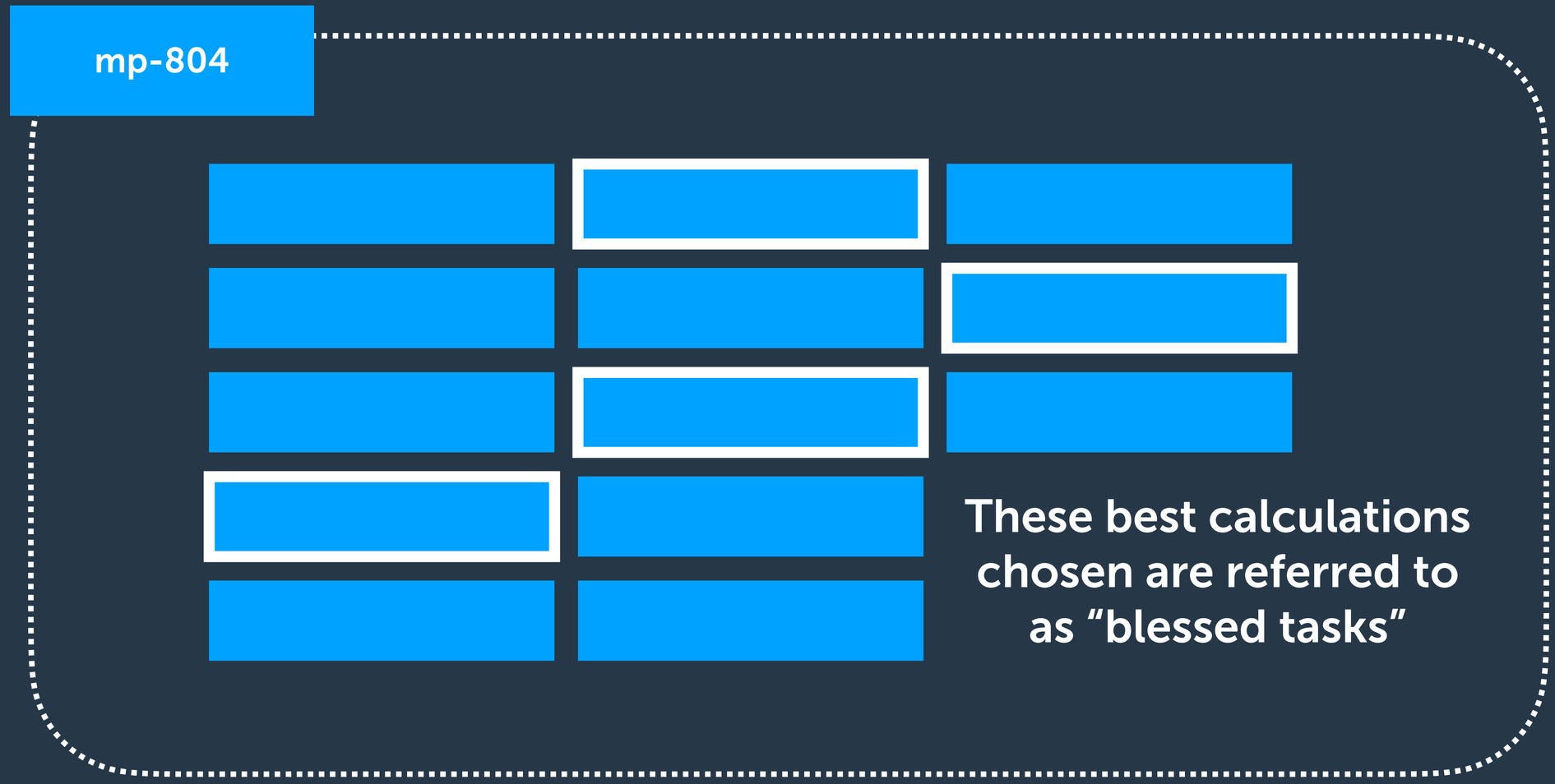


Task IDs are used to reference specific individual calculations

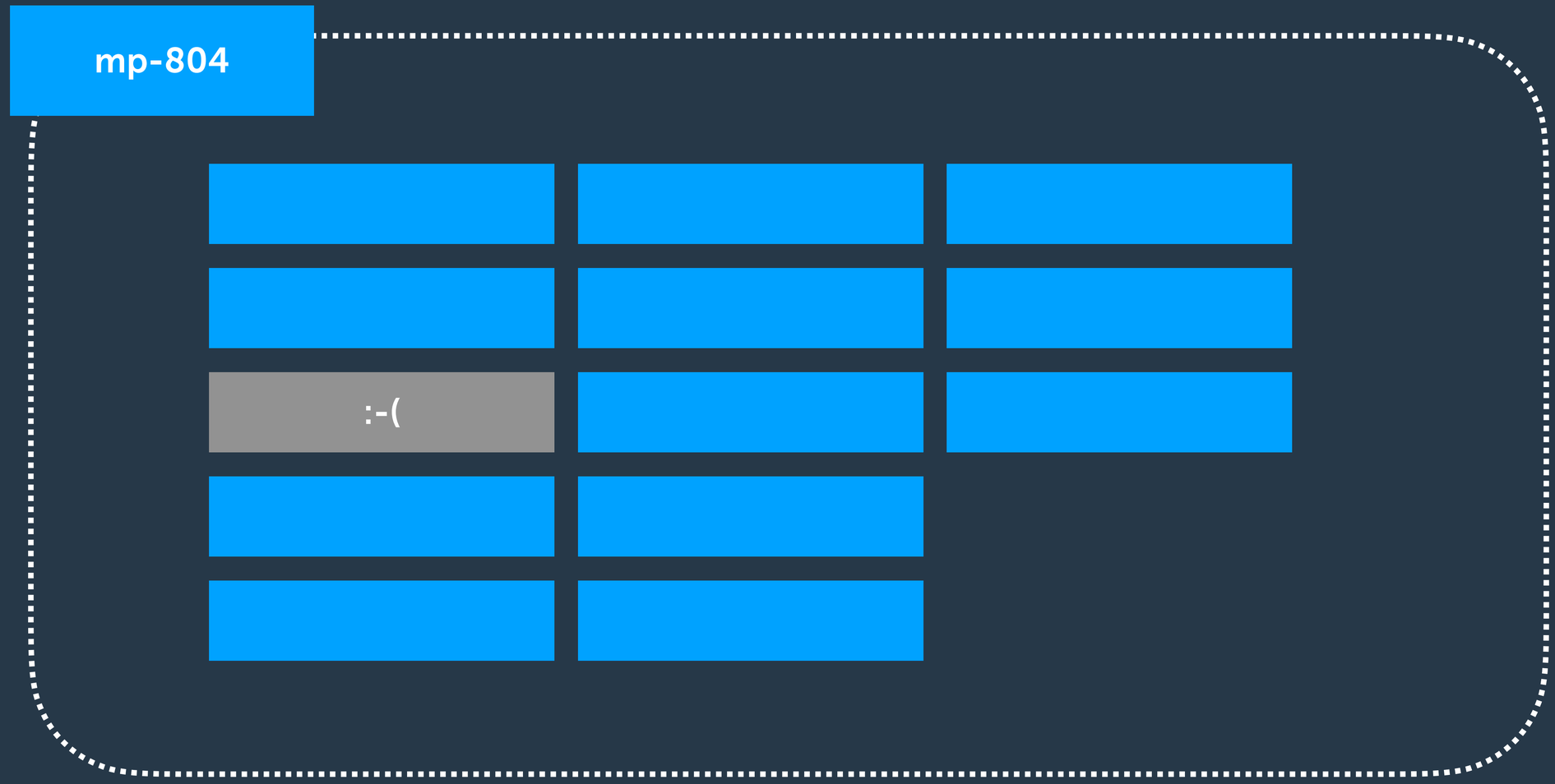
Often, we will add new calculation tasks for existing materials as we improve our methods



When we have multiple tasks for the same property, we will choose the “best” one to report on the Materials Project website

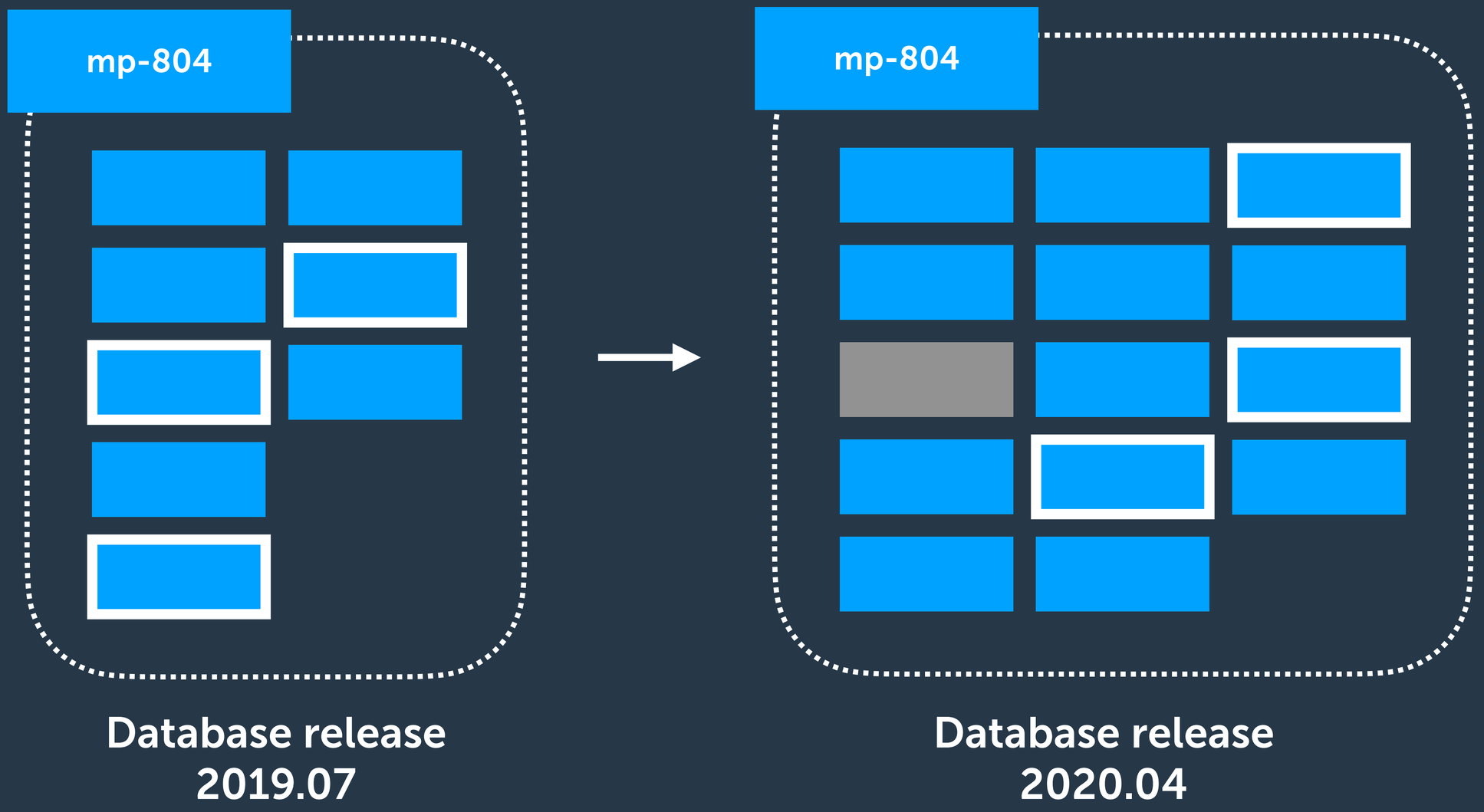


When we detect an error in a task or a task no longer meets our quality standards, we deprecate it



The data from a deprecated task still remains available, but is no longer shown by default on the website

Materials Project updates its data during a database release



Our implicit user contract

Task ID: This is the raw calculation data and will always remain constant and accessible wherever possible. The final energy of a given task will never change, the band gap for a given task will never change.

Material ID: Combined, aggregated data reported on the MP website under its Material ID can and *will* change with new database releases. This happens as we deprecate old tasks, introduce new tasks, and change which tasks we "bless" as a result.

When citing Materials Project data, make sure to include the database version!

We are improving this process as time goes on. We are committed to FAIR data principles and our data is available through an open CC-BY-4.0 license.

Website Demo