A couple of resources that might be helpful for your project:

1. [**TM23 Data Set**: This data set examines the accuracy of machine-learned force fields (MLFFs) for bulk solid and liquid phases of d-block elements1](https://arxiv.org/abs/2302.12993). [It contrasts the performance of force, energy, and stress predictions across the transition metals for two leading MLFF models1](https://arxiv.org/abs/2302.12993). You can find more details about this dataset [here](https://arxiv.org/abs/2302.12993).
2. [**Quantum-Machine.org Datasets**: This website provides a variety of datasets for quantum machine learning](https://arxiv.org/abs/2302.12993)[2](http://quantum-machine.org/datasets/). [One of them is the QM7Dataset, which is a subset of GDB-13 (a database of nearly 1 billion stable and synthetically accessible organic molecules) composed of all molecules of up to 23 atoms2](http://quantum-machine.org/datasets/). You can find more details about this dataset [here](http://quantum-machine.org/datasets/).
3. [**Accurate global machine learning force fields for molecules with hundreds of atoms**: This paper might provide some insights into your research3](https://www.science.org/doi/pdf/10.1126/sciadv.adf0873?download=true). You can find more details about this paper [here](https://www.science.org/doi/pdf/10.1126/sciadv.adf0873?download=true).
4. [**A collection of forcefield precursors for metal–organic frameworks**: This paper provides a list of computed forcefield precursors4](https://pubs.rsc.org/en/content/articlehtml/2019/ra/c9ra07327b). You can find more details about this paper [here](https://pubs.rsc.org/en/content/articlehtml/2019/ra/c9ra07327b).

Remember, when searching for datasets, it’s important to consider the size, quality, and relevance of the data to your specific project. Good luck with your thesis!