

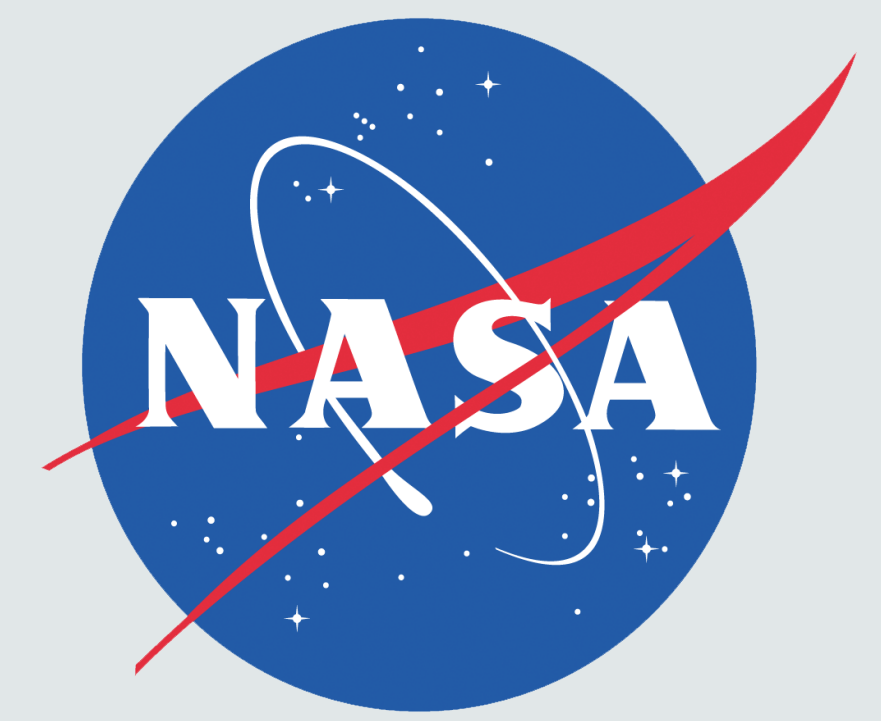
Harvesting NASA's Common Metadata Repository

Dana Shum¹

dshum@raytheon.com

Chris Durbin¹ | James Norton² | Andrew Mitchell³

IN51A-0003

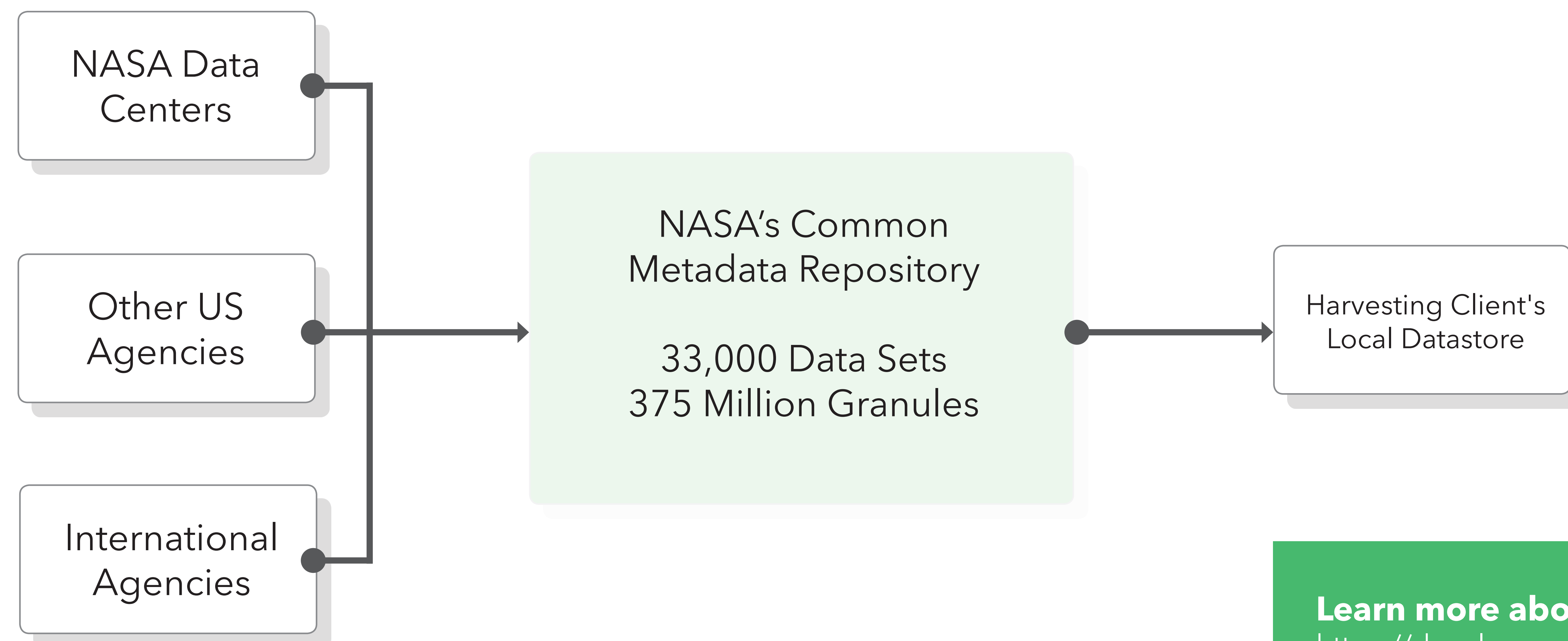


NASA's Common Metadata Repository (CMR) contains a vast amount of Earth Science metadata representing both US and International datasets.

While the CMR's robust searching capabilities make it ideal for clients to use directly, the CMR now also supports the ability to harvest CMR's holdings. This enables clients to store the CMR's holdings locally and potentially combine the CMR's holdings with other non-CMR data.

The CMR Harvesting feature, added in 2017, supports the following types of requests:

- Return all datasets (collections)
- Return all granules for a given dataset
- Return datasets or granules modified after a given date
- Return datasets or granules added after a given date
- Return datasets or granules deleted after a given date



Learn more about harvesting:
<https://developer.earthdata.nasa.gov/cmr/harvesting-best-practices>

Technical Challenges

The CMR's harvesting functionality was implemented to ensure that the following criteria are met:

- Large result sets can be retrieved (>1 million results)
- While iterating through result sets, the harvesting request results remain consistent
- Performance of other queries in the system are unaffected by harvesting requests

