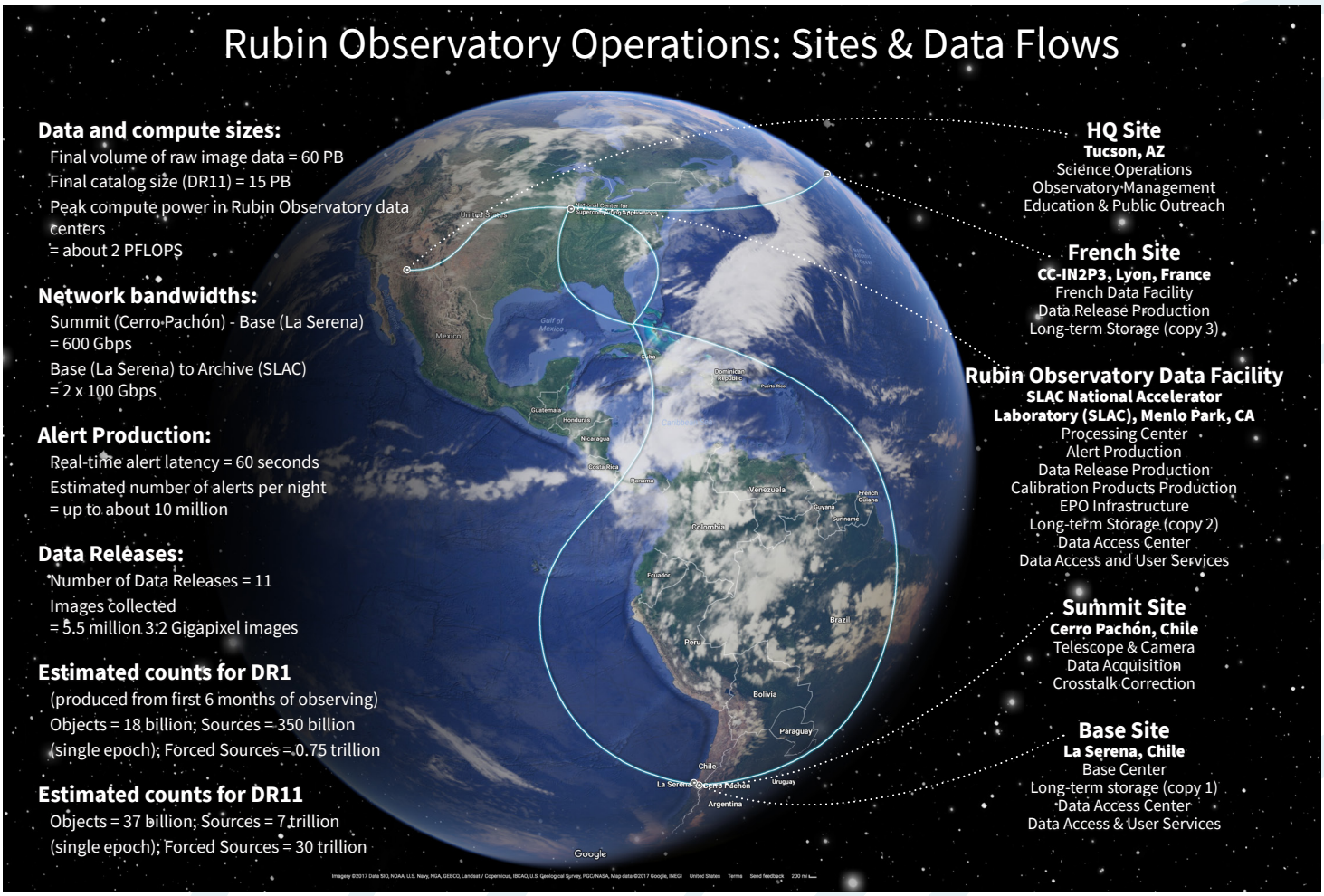
RUBIN OBSERVATORY & BIG DATA

The term “Big Data” refers not only to the volume of data, but the complexity of data processing required to extract knowledge from it. The old method of gathering data to answer a question or validate a hypothesis is giving way to data-driven exploration and discovery using massive datasets collected in survey mode. Software can explore the high dimensionality of Big Data to discover connections and relationships not known ahead of time; the data itself drives the discovery.

Rubin Observatory is not only a telescope, it is a database—a Big Data project tuned to facilitate data-driven explorations of the most fundamental questions of the Universe. Rubin Observatory will survey the visible night sky every night for ten years, building a 500 PB database of images and a 15 PB catalog of text data describing properties of nearly 40 billion individual stars and galaxies.



Cutting-edge computer applications will be used to hold the data and mine it for scientific discoveries. The data management system being developed must process the nightly alert data, 20,000 expected alerts per minute, in near real time, and construct annual data releases (DRs) at the petabyte scale. Each DR processes image data collected from the start of the survey. Rubin Observatory focuses on providing well-calibrated data to the community through high-performance computing interfaces. The software development team consists of more than 100 people working in six different sites across the US, with further contributions from IN2P3 in France, to develop an integrated software set that will help realize the Rubin Observatory science goals.

Analysis of Rubin Observatory database will address four primary science areas:

Probing Dark Matter and Dark Energy | Taking an Inventory of the Solar System | Exploring the Changing Sky | Mapping the Milky Way

D-14554 12/21