Goals
- To provide exascale ready computational and data services that will accelerate innovation
- To validate the services in real-world settings, both in scientific research and in industry pilot deployments:
  - Square Kilometre Array – a large radiotelescope project
  - medical informatics
  - airline revenue management
  - open data for global disaster risk reduction
  - agricultural analysis based on Copernicus data

Extreme Large Computing Services

Survey of interactive execution environments
- Focus on:
  - integration of scripting notebooks with HPC infrastructures to support building extreme large computing services
  - extension mechanisms required to add support specific to exascale processing of large data sets
  - ability to mix multiple languages in one document
  - integration with cloud infrastructures

<table>
<thead>
<tr>
<th>Name</th>
<th>Large data set support</th>
<th>Integration with Cloud/HPC infrastructures</th>
<th>Extension mechanisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Notebook</td>
<td>using additional custom libraries (e.g. for Apache SPARK)</td>
<td>using custom libraries communicating with HPC queuing systems (e.g. SLURM)</td>
<td>It is possible to develop custom engines for languages which are not natively supported.</td>
</tr>
<tr>
<td>DataBricks</td>
<td>the whole platform is based on Apache SPARK</td>
<td>Available only on Amazon Web Services or Microsoft Azure</td>
<td>almost none</td>
</tr>
<tr>
<td>Beaker</td>
<td>using additional custom libraries</td>
<td>no specific support for HPC; Docker version available</td>
<td>Users can add Beaker support for unsupported languages via a dedicated API.</td>
</tr>
<tr>
<td>Jupyter</td>
<td>using additional custom libraries</td>
<td>no mature solution for HPC; Docker version available</td>
<td>Additional languages can be supported by writing a new Jupyter kernel.</td>
</tr>
<tr>
<td>Cloud Datalab</td>
<td>support for Google data services (e.g. BigQuery, Cloud Machine Learning Engine, etc.)</td>
<td>restricted to the Google Cloud platform</td>
<td>limited</td>
</tr>
<tr>
<td>Zeppelin</td>
<td>native support for Apache Spark</td>
<td>can be run on HPC using connection to the YARN cluster</td>
<td>support for additional languages can be added</td>
</tr>
</tbody>
</table>

Summary
- DataBricks and Cloud Datalab must be run on specific cloud resources
- Zeppelin and DataBricks are based on Apache SPARK, which potentially limits their usage to that platform
- R Notebooks seems promising; however, some important features are only available with a commercial version of RStudio
- BeakerX (successor to Beaker) and Cloud Data are based on the Jupyter solution
- Jupyter seems to be a suitable base for developing extreme large computing environments

References
2. Beaker Notebook webpage http://beakernotebook.com/features
4. Datalab webpage https://cloud.google.com/datalab/
5. Jupyter webpage http://jupyter.org/
6. Rstudio webpage https://www.rstudio.com