

## Orphan Drug Development Guidebook

### Building Block I407

This document defines the content of the Building Block created for each identified tool, incentives, initiative or practice introduced by public bodies or used by developers to expedite drug development in Rare Diseases (RDs).

ITEM	DESCRIPTION
Building Block (BB) Title	Data Mining
References	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6794821/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6794821/</a>
Description	<p>The understanding of rare diseases remains limited at a time when the capacity to generate data continues to grow. These unprecedentedly large amounts of data – from rare to non-rare to common diseases – have challenged researchers trying to make sense of it. Meanwhile, data-sharing initiatives also opens up access to new types of data including patient records and other real-world data. These data are ripe for analyses using big data techniques, including computational models that unveil molecular mechanisms and similarities among clinical phenotypes, predict compound-ligand interactions, perform high-throughput screening of molecules against cell lines and network-based in silico drug efficacy screening, and data mining for potential therapeutic targets based on existing knowledge. At the point of convergence of several academic research fields (e.g., applied mathematics, computer science, artificial intelligence, statistics and machine learning), data mining takes advantage of the potential to carry out novel multi-dimensional analytics to connect data on diseases, mechanisms, proteins, and drugs.</p> <p>Pieced together, data mining methods enable the discovery of new or the repurposing of previously known pharmaceutical compounds in the development of treatments for new indications.</p>
Category	Development Opportunity Building Block

Geographical scope	International
Availability	Applicants developing medicines for rare and non-rare diseases.
Scope of use	To discover new connections and ideas, based on already available data combining different types/kinds of data and extract the most available information.
Stakeholders	<ul style="list-style-type: none"> <li>• Researchers (public and private)</li> <li>• Data experts</li> </ul>
Enablers / Requirements	N/A
Output	New knowledge based on already available datasets.
Best time to apply and time window	The tool has its best use especially in the early phases of drug development, to gather the maximum of information.
Expert tips	<p>If you do intend to use data mining strategies, make sure you have several robust data sets, including different data sources, such as biological data, patient health record data, and other real-world data.</p> <p>PROs:</p> <p>Allows you to systematically discover new correlations that you might have missed otherwise.</p> <p>CONs:</p> <p>Sufficient (freely accessible) data is needed if you want to be able to find new information</p>