

## Orphan Drug Development Guidebook Task Force

### Galaxy Guide for drug development

#### Building Block FACT SHEET FORM content

This document defines the content of the FACT SHEET to be created for each identified tool, incentives, initiative or practice (the Building Block) 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/pubmed/28129139">https://www.ncbi.nlm.nih.gov/pubmed/28129139</a>
Description	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 <i>in silico</i> 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.
Relevance to rare disease drug development	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.
Category	Development opportunity
Availability	Data mining is a tool, accessible for use in both the public and

ITEM	DESCRIPTION
	private domain, for rare and common diseases, or even outside of the disease context.
Geographical scope	Globally
Scope of use	To combine different types and kinds of data, and extract most available information from this.
Subject	Researchers
Enablers/ Requirements	NA
Output	New knowledge based on already available datasets.
Actors and Stakeholders	Researchers (public and private); data experts
Use	To discover new connections and ideas, based on already available data.
PROs/ advantages	Allows you to systematically discover new correlations that you might have missed otherwise.
CONs/ risks	Sufficient (freely accessible) data is needed if you want to be able to find new information
Best time to apply	Especially in the early phases of drug development, to gather the maximum of information
Duration	Several weeks
Cost	No costs
Practical tips	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.

