



## Societal Challenge nr.1: Life Sciences & Health

**Big Data Focus area:** Large-scale heterogeneous pharma-research data linking & integration

**BDE2020 domain partners:**



**Selected Key Data assets:** ACD Labs / ChemSpider, ChEBI, ChEMBL, ConceptWiki, DrugBank, ENZYME, Gene Ontology, GO Annotation, SwissProt, WikiPathways

### Pilot 1:

Duplicate [Open PHACTS functionality](#) on the BDE infrastructure using Open Source solutions

Reasons:

- Deployment possible in-house on an Intranet which improves security
- Apply the system to domains (e.g. Agriculture)
- Using extra BDE components (e.g. logging, analysis) for extra functionality.

BDE infrastructure functionality

- Large scale data extraction and integration processing from external data sources (tables, figures texts)
- Analysis batch jobs for generating statistical data
- Rich query support combining various parameters (e.g. location, geno/fenotypes, publications, soil data)
- Various front ends similar to PubMed

