



---

## Data Infrastructure for Chemical Safety

Across Europe chemical safety is receiving growing attention, and there is an increasing demand for the use of non-animal tests to study toxicity in human cells. In response, the EC has funded many projects that investigate the use of data-dense genomics technologies to predict toxicity; however, there is currently no infrastructure for capturing the data produced by these projects in a standardized, harmonized and sustainable manner.

In the foreseeable future, large data sets will be collected on the chemical, physical and toxicogenomic properties of chemicals, through application of high density chemical analysis, multiple omics technologies, live cellular imaging, and so on. Without a unifying infrastructure for storing this data, there is a risk that valuable data – key to enabling innovative breakthroughs for the toxicogenomics research community – may evaporate.

[diXa](#) is a new initiative designed to fill this gap. diXa aims to develop a web-based, open access and sustainable e-infrastructure for storing and searching data sets produced by past, current and future EU research projects that target non-animal chemical safety tests. These data sets will be linked to other databases for chemical, physical and/or toxicological information, as well as to databases on molecular medicine, thus crossing the traditional borders between scientific disciplines and reaching out to other research communities. In addition, diXa will offer services and procedures for data generation, harmonization and standardization, as well as customized tools and techniques for advanced statistics and modelling.

The project includes two major phases: phase one focuses on data, including data organization, formats, procedures, and metadata specifications. Phase two focuses on computational challenges, for example, the development of pattern recognition services for use once the data is stored in proper repositories and can be accessed.

diXa is collaborating with several other initiatives to build this infrastructure, in particular with the European Bioinformatics Institute (EBI) in Cambridge. diXa and EUDAT are holding ongoing discussions to identify and optimise opportunities for collaboration between the two projects: EUDAT provides an umbrella for globalizing diXa's efforts. As we are looking for ways to carry out compute-intensive jobs on our data sets, we need to be able to capture and exploit the increasing complexity and speed of omics-based data generation and next-gen sequencing.

The work performed in collaboration with diXa was concluded in March 2015.

[Read more](#)