

diXa: A new data infrastructure for chemical safety

Jos Kleinjans introduces the Data Infrastructure for Chemical Safety (diXa) project and explains what diXa hopes to gain from partnering with EUDAT.

FEATURE | SEPTEMBER 12, 2012 | BY JOS KLEINJANS

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, according to a research paper in the journal *Science*. The European Commission 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.

Soon, large data sets will be collected on the chemical, physical, and toxico-genomic properties of chemicals, through application of high-density chemical analysis, multiple 'omics' (for example genomics or proteomics) technologies, and live cellular imaging. Without a unifying infrastructure for storing this data, there is a risk that valuable data – key to innovative breakthroughs for the toxico-genomics research community – may evaporate.

The 'Data Infrastructure for Chemical Safety' (diXa) project is a new initiative designed to fill this gap. It aims to develop a web-based, open-access, and sustainable e-infrastructure for storing and searching data sets produced by past (i.e. EC FP6 projects), current (i.e. FP7 projects), and future EC research projects that target non-animal chemical safety tests.

The project includes two major phases: the first focuses on upgrades to the existing infrastructure and its data, including data organization, formats, procedures, and metadata specifications. The second, running simultaneously, focuses on computational challenges, e.g. the development of pattern-recognition services for use once the data is stored in proper repositories and can be accessed. These phases are due to finish in 2013.

The data sets will be linked to other databases of chemical, physical, and toxicological information, and to databases on molecular medicine, thus crossing the traditional borders between scientific disciplines and reaching out to other research communities.

The initial users of diXa will be toxicologists. They'll be able to exploit data-rich genomics technologies for better understanding of toxic mechanisms and predicting the safety of chemicals in human health. Other research communities such as systems biologists, bioinformaticians, biostatisticians, mathematicians, and computer scientists will also use the service.

In addition, diXa will offer services and procedures for data generation, harmonization, and standardization. One standard that will be used to bring a common file format to represent experimental metadata is the [ISA software tool box](#). This is a software suite designed to manage omics data and its associated metadata by leveraging common elements while keeping data files external in their native or community-specific formats.

Globalizing the infrastructure

DiXa is collaborating with several other initiatives to build its infrastructure, in particular with the [European Bioinformatics Institute](#) (EBI), in Cambridge. EBI already has an existing platform called ArrayExpress for uploading ISA-formatted datasets for toxico-genomics, which may run on grid computing.

To help analyze and manage diXa's oncoming torrent of data, one solution is to store its data close to powerful computer servers, such as those available through EUDAT. The [EUDAT project](#) is holding ongoing discussions with diXa to identify and optimize opportunities for collaboration.

"DiXa is very relevant to EUDAT and bears some similarity to what EUDAT is trying to achieve itself at a pan-European level. EUDAT is willing to support this initiative by providing a robust storage and computing infrastructure to complement the solutions being designed. DiXa's trans-disciplinary approach is also very interesting for us and we look forward to fruitful discussions and collaboration between diXa and the other research communities involved in EUDAT," said Damien Lecarpentier of the EUDAT project.

EUDAT will provide an umbrella for globalizing diXa's efforts. The [iCORDI project](#), which is under the EUDAT umbrella, will enable diXa to explore joint efforts directly with US data infrastructures.

On 25 May 2012, diXa and DG INFSO organized a high level workshop for experts from multiple disciplines in academia and industry. DiXa will present the outcomes of this workshop at the first [EUDAT conference in October 2012](#).

Average:

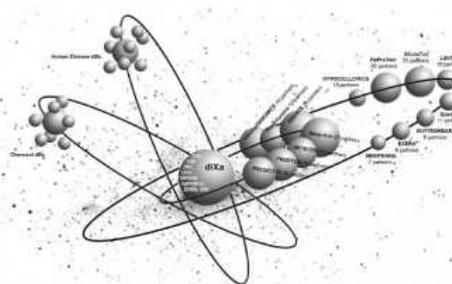
Your rating: None Average: 4 (5 votes)

About the Author »

Jos Kleinjans

University of Maastricht, Netherlands, and diXa Project Coordinator.

RELATED TERMS: [DG INFSO](#) [diXa](#) [EBI](#) [EUDAT](#) [Europe](#) [FP6](#) [FP7](#) [grid computing](#) [iCORDI](#) [ISA](#)
[biology](#) [North America](#) [chemistry](#) [data management systems](#) [health and medicine](#)



The diXa universe: Presentation of the diXa network. Its partners are within the core of the network and contributing FP6/FP7 toxico-genomics projects are in the immediate surroundings. The distant FP6/FP7 projects pursue development of non-animal-based chemical safety tests, but without exploiting 'omics' technologies. The intersecting orbits show the research communities of clinical sciences and computational chemistry/chemo-informatics. Image courtesy diXa.

Comments

ADD NEW COMMENT

Post new comment

Subject:

Comment: *

[Input format](#)

By submitting this form, you accept the [Mollom privacy policy](#).