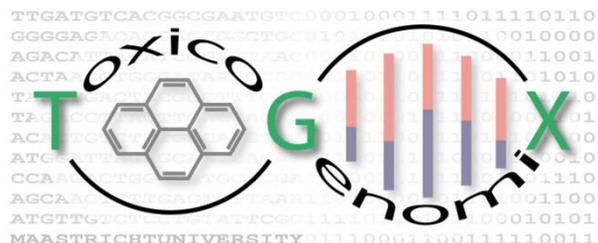


# diXA - a **d**ata **i**nfrastructure for **c**hemical **s**afety

Jos Kleinjans  
Dept of Toxicogenomics  
Maastricht University



# Current protocol for chemical safety testing

- Short-Term Tests for Genetic Toxicity
  - **Bacterial Reverse Mutation Test**
  - ***In vitro* Mammalian Chromosomal Aberration Test**
  - ***In vitro* Mouse Lymphoma TK +/- Gene Mutation Assay**
  - ***In vivo* Mammalian Erythrocyte Micronucleus Test**
- Acute Oral Toxicity Tests
- Short Term Toxicity Studies
  - Short-Term Toxicity Studies with Rodents (3 dose levels, 2-4 weeks)
- Subchronic Toxicity Studies
  - Subchronic Toxicity Studies with Rodents (3 dose levels, 90 days)
- Carcinogenicity Studies with Rodents
- Combined Chronic Toxicity/Carcinogenicity Studies with Rodents
- *In Utero* Exposure Phase for Addition to Carcinogenicity Studies with Rodents
- Reproduction and Developmental Toxicity Studies
  - Multi-generation Reproduction Studies (3 dose levels, life time)
  - Stand alone/multi-generation Developmental Toxicity Studies
- Neurotoxicity Studies
- Metabolism and Pharmacokinetic Studies
- Immunotoxicity Studies

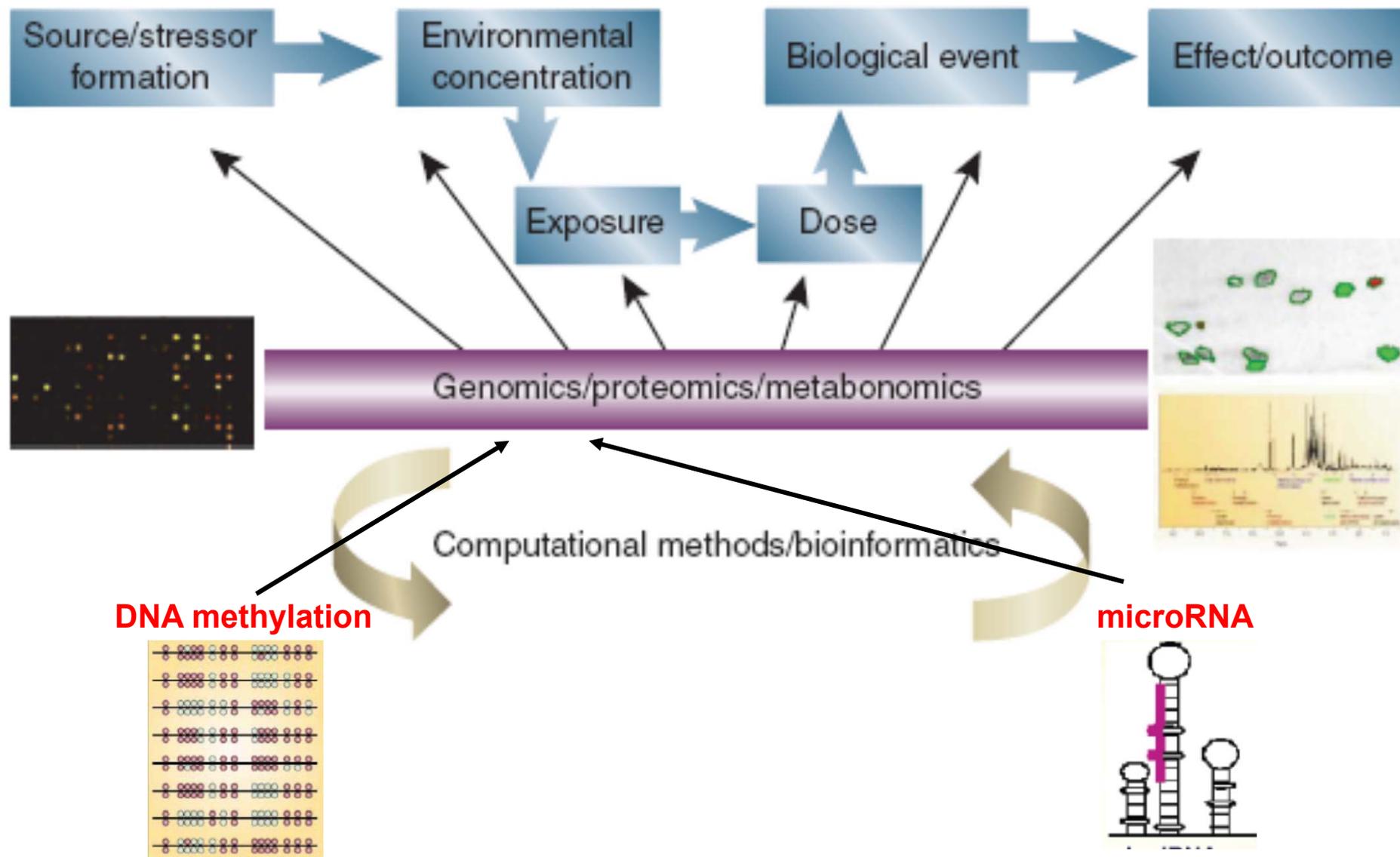
## Increasing demands on chemical risk assessment:

- High failure rate of new drug candidates due to unmanageable toxicity, accounting for approximately 30% of this attrition
- The EU REACH program on industrial chemicals
  - **R**egistration, **E**valuation and **A**uthorization of **C**hemicals:
  - Existing and new substances should in the future be subject to the same procedure under a *single system*.
  - Large amounts of additional tests required before 2018
    - 30,000 existing chemicals already placed on the market since before 1981 and sold at > 1 tonne per year
- EU-wide ban on animal use in cosmetics development
  - Council Directive 92/32/EEC (7<sup>th</sup> Amendment): testing ban prohibiting animal testing of cosmetic products in the EU.
- Future EU regulations on food chemicals

## The TeGenero case (*NJEM, 2006: 354: 1869-1871*)

- At 8 a.m. on Monday, March 13, 2006, eight healthy young men entered a trial of a drug under development by the small German immunotherapeutics company TeGenero. Six of the volunteers were assigned to receive active drug, and two were to receive placebo.
- The six volunteers were to be the first humans to receive TGN1412, a humanized monoclonal antibody designed as an agonist of the CD28 receptor on T lymphocytes, which stimulates the production and activation of T lymphocytes.
- It was hoped that this product would benefit patients with B-cell chronic lymphocytic leukemia or autoimmune diseases such as multiple sclerosis or rheumatoid arthritis.
- However, after receiving injections of TGN1412, the six volunteers became desperately ill, had multiple-organ failure, and were transferred to an intensive care unit with what has been described as a cytokine release syndrome.
- Preclinical testing, including tests in rabbits and monkeys that used doses up to 500 times as high as the doses received by the first group of volunteers, reportedly showed no signs of toxicity.

# Capturing the complexity: integration of transcriptomics, proteomics, metabonomics with epigenetics and $\mu$ RNA and bioinformatics applying human cellular models



# Big data in toxicogenomics

- Acquisition of the knowledge to develop high-throughput testing assays would involve the discovery of **toxicity pathways and networks** from vast amounts of data... **Central repositories for -omics data** ... exist to a small extent ... The scale of data storage and access envisioned by the committee is much larger - *Toxicity Testing in the 21st Century (2007)*
- The vision is a **scientific community that does not waste resources** on recreating data that have already been produced, in particular if public money has helped to collect those data in the first place. Scientists should be able to concentrate on the best ways to make *use* of data. Data become an infrastructure that scientists can use on their way to new frontiers – *A Digital Agenda for Europe (2010)*
- But **digital information is inherently fragile and often at risk of loss**. Access to valuable digital materials tomorrow depends upon preservation actions taken today; and, over time, access depends on ongoing and efficient allocation of resources to preservation - *Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010)*



## Main objectives of EU/FP7's diXa

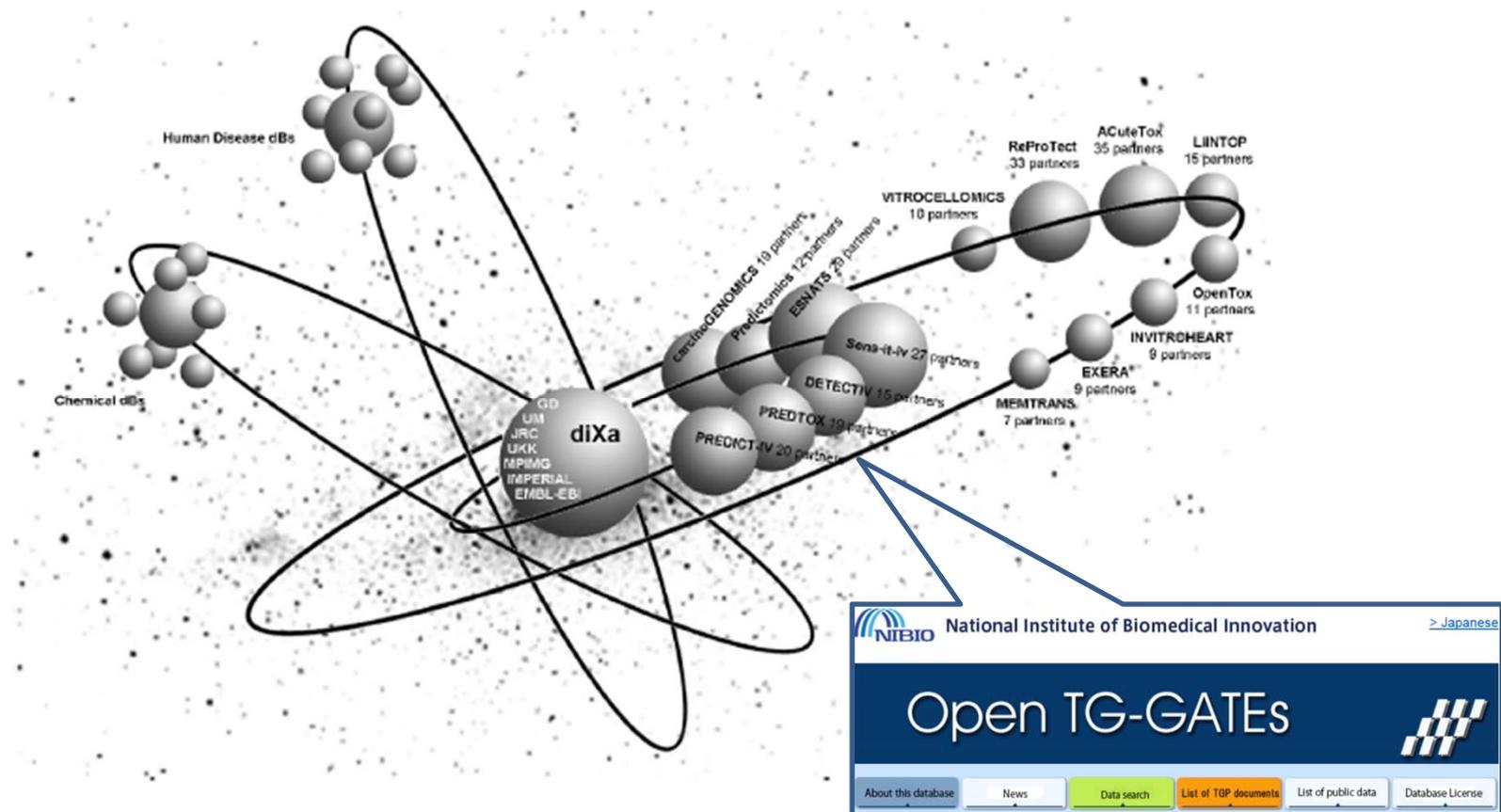
- To further develop and adopt a robust and sustainable service infrastructure (e.g. data infrastructure and e-science environment) for harbouring multiplexed data sets as produced by past, current and future EU research projects on developing non-animal tests for predicting chemical safety as conducted by the research community of toxicogenomics
- To link this with other research communities maintaining globally available chemical/toxicological data bases and data bases on molecular data of human disease.

## diXa's research aims

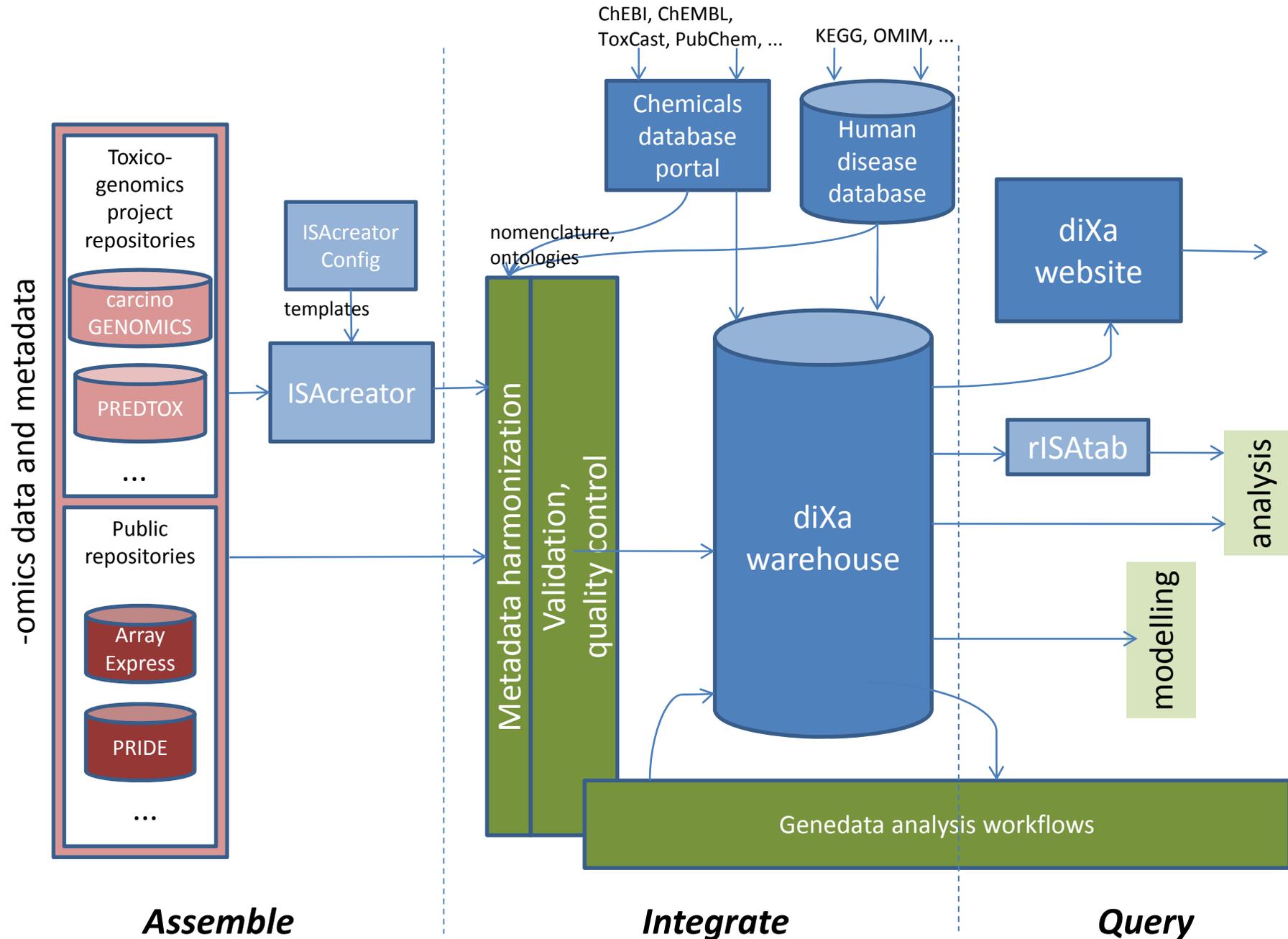
- The main scientific concept which has initiated the diXa proposal, is
  - to create a large public data infrastructure of genomic signatures of drugs, industrial chemicals and cosmetics, and
  - to develop pattern-matching bioinformatics and biostatistics tools to detect similarities among these signatures,
  - in order to describe all biological states induced with a chemical exposure, in terms of genomic signatures relevant for the human situation *in vivo*.

# Creating diXa's data infrastructure

- \* by integrating TGX data from FP6/FP7 projects
- \* by linking with chemical/mol. medicine data bases
- \* thus enabling cross-platform, cross-study integrations



# Full walk-through from data sources to data consumers



# Toward interoperable bioscience data

Susanna-Assunta Sansone<sup>1,39</sup>, Philippe Rocca-Serra<sup>1,39</sup>, Dawn Field<sup>2</sup>, Eamonn Maguire<sup>1</sup>, Chris Taylor<sup>2,3</sup>, Oliver Hofmann<sup>4</sup>, Hong Fang<sup>5</sup>, Steffen Neumann<sup>6</sup>, Weida Tong<sup>7</sup>, Linda Amaral-Zettler<sup>8</sup>, Kimberly Begley<sup>4,9</sup>, Tim Booth<sup>2</sup>, Lydie Bougueleret<sup>10</sup>, Gully Burns<sup>11</sup>, Brad Chapman<sup>4</sup>, Tim Clark<sup>12,13</sup>, Lee-Ann Coleman<sup>14</sup>, Jay Copeland<sup>15</sup>, Sudeshna Das<sup>12,13</sup>, Antoine de Daruvar<sup>16,17</sup>, Paula de Matos<sup>3</sup>, Ian Dix<sup>18</sup>, Scott Edmunds<sup>19</sup>, Chris T Evelo<sup>20,21</sup>, Mark J Forster<sup>22</sup>, Pascale Gaudet<sup>23,24</sup>, Jack Gilbert<sup>25</sup>, Carole Goble<sup>26</sup>, Julian L Griffin<sup>27,28</sup>, Daniel Jacob<sup>17,29</sup>, Jos Kleinjans<sup>30</sup>, Lee Harland<sup>31</sup>, Kenneth Haug<sup>3</sup>, Henning Hermjakob<sup>3</sup>, Shannan J Ho Sui<sup>4</sup>, Alain Laederach<sup>32</sup>, Shaoguang Liang<sup>19</sup>, Stephen Marshall<sup>33</sup>, Annette McGrath<sup>34</sup>, Emily Merrill<sup>13</sup>, Dorothy Reilly<sup>33</sup>, Magali Roux<sup>35,36</sup>, Caroline E Shamu<sup>15</sup>, Catherine A Shang<sup>37</sup>, Christoph Steinbeck<sup>3</sup>, Anne Trefethen<sup>1</sup>, Bryn Williams-Jones<sup>31</sup>, Katherine Wolstencroft<sup>26</sup>, Ioannis Xenarios<sup>10,38</sup> & Winston Hide<sup>4</sup>

To make full use of research data, the bioscience community needs to adopt technologies and reward mechanisms that support interoperability and promote the growth of an open 'data commoning' culture. Here we describe the prerequisites for data commoning and present an established and growing ecosystem of solutions using the shared 'Investigation-Study-Assay' framework to support that vision.

To tackle complex scientific questions, experimental datasets from different sources often need to be harmonized in regard to structure, formatting and annotation so as to open their content to (integrative) analysis. Vast swathes of bioscience data remain locked in esoteric formats, are described using nonstandard terminology, lack sufficient contextual information or simply are never shared due to the perceived cost or futility of the exercise. This loss of value continues to engender standardization initia-

service providers and circumvents many of the problems caused by data diversity. The same framework enables researchers, bioinformaticians and data managers to operate within an open data commons.

## From reusable data to reproducible research

Shared, annotated research data and methods offer new discovery opportunities and prevent unnecessary repetition of work. Although

through the provision of independent databases, tools and curators, driven by advocates of the sharing of both pre- and post-publication data<sup>7,8</sup>. To build an interoperable open data ecosystem will require leveraging all of these positive efforts and further increasing community buy-in.

## Time to leap outside the box

Overall, most stakeholder groups accept the principles of data sharing, but in practice,

HOME ABOUT ICORDI ECOSYSTEM NEWS ROOM EVENTS

Email

Password

Remember me

Register | [Forgot Password](#)



*iCORDI - The premier global forum charting, demonstrating and driving convergence between emerging global data infrastructures*

High Level Strategic Forum    Research Data Alliance

Subscribe to the iCORDI Newsletter

iCORDI All Social

My iCORDI

Latest from iCORDI

*(KE)* 15.04.2013

iCORDI Ecosystem: user communities & prototypes

DATA-INTENSIVE ASTRONOMY



EARTH SCIENCES



OPEN DATA & INTEROPERABILITY



CHEMICAL SAFETY



**iCORDI focuses on coordinating a series of cross-infrastructure experiments on global interoperability with a selected group of projects and communities. Each prototype addresses a specific community-driven use case identifying best-of-breed solutions and the remaining challenges.**

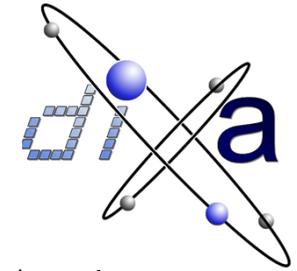


# Prototype: Towards global connections in chemical safety



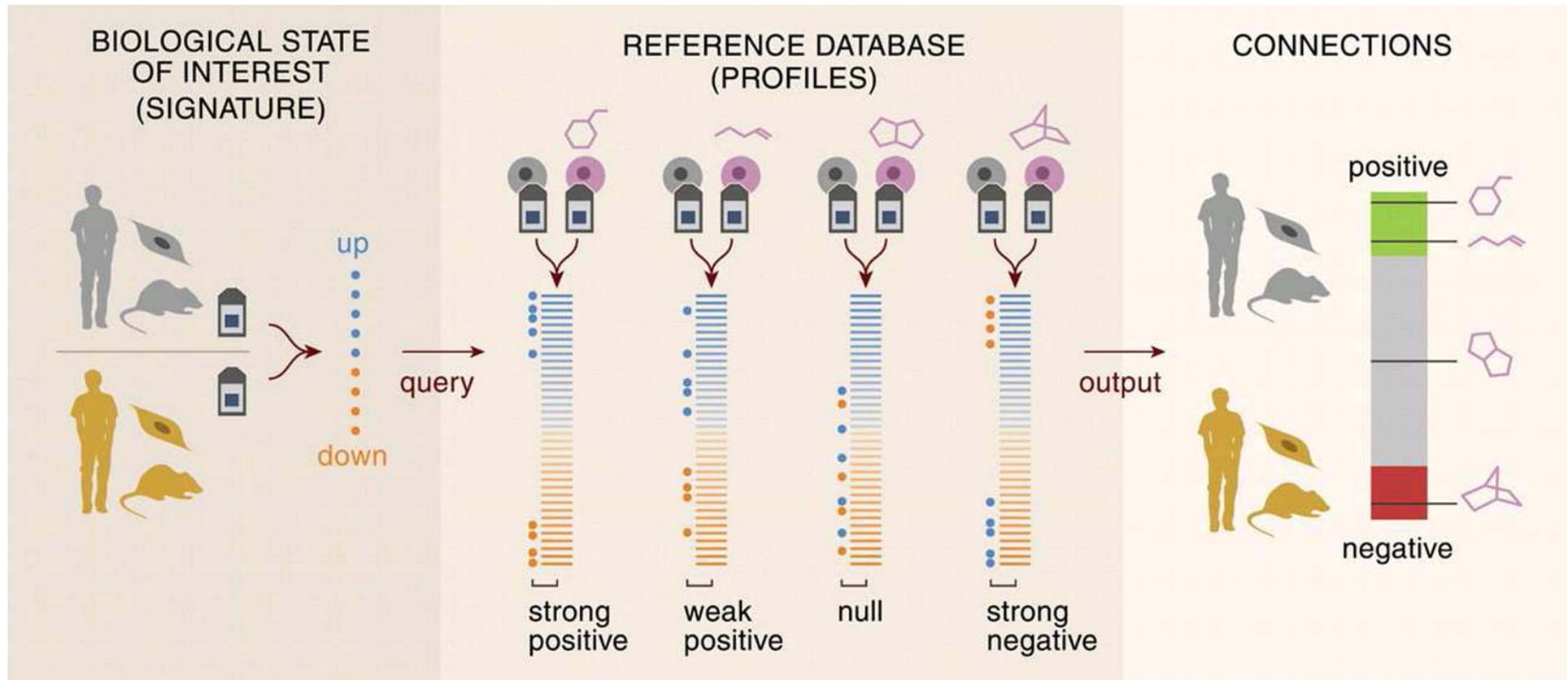
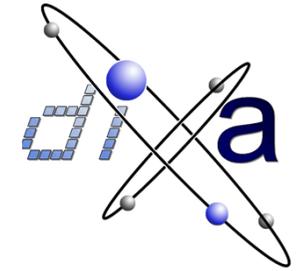
- What specific challenges will arise in linking to additional databases held by US researchers?
- Will diXa's service-based mechanisms also work for US dBs, or must they be adapted?
- What steps can we take now towards common metadata or data scheme patterns for global chemical safety databases?

# Overview of CP-CSA activities



- Networking activities
  - fully functional, web-based, openly accessible and sustainable e-infrastructure for capturing toxicogenomics data from relevant EU FP6/FP7 projects
  - linking to available data bases holding chemico/physico/toxicological information
  - linking to data bases on molecular medicine
- Servicing support from this Collaborative Data Infrastructure
  - clear communication channels with the TGX research community
  - deliver commonly agreed core service support
    - by providing SOPs for seamless data sharing
    - by offering quality assessments and newly developed tools and techniques for data management
    - by offering access to the diXA infrastructure for toxicogenomics scientists
  - all supported by hands-on training.
- Joint research initiative
  - cross-platform integrative statistical analyses
  - cross-study meta-analyses
  - systems modeling for predicting chemical safety as alternative to animal tests
- Thus contributing to the vision expressed by Riding the Wave and Europe's Digital Agenda

# Visualization of the potential use of diXa's end-product

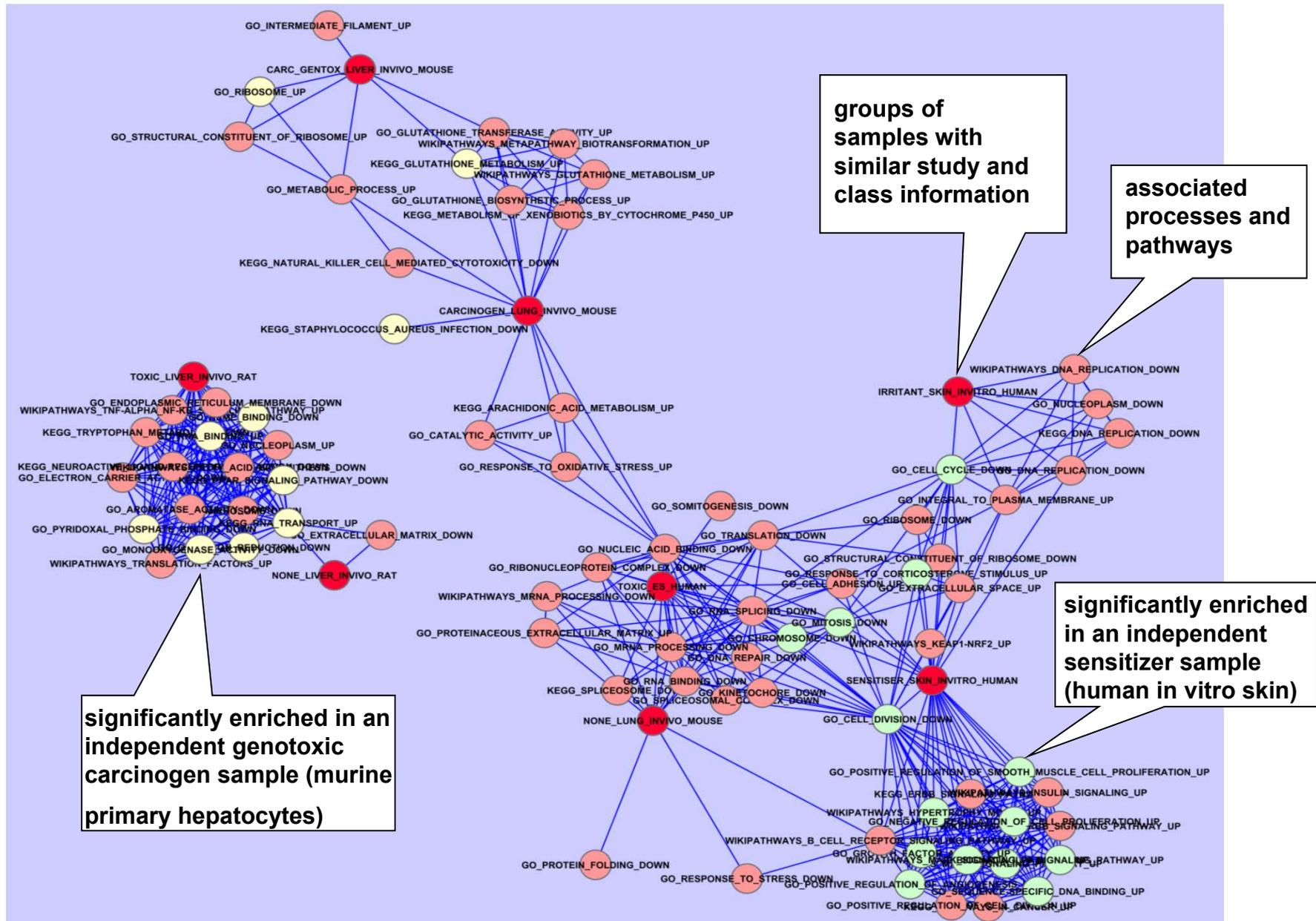


*J. Lamb et al. The Connectivity Map: using gene-expression signatures ...Science. 2006*



**Pilot on perturbed pathways in liver toxicity**

# NTC-wide associations between toxicity class, compounds and processes



# Future perspective: Personalized Safety Is A Big Challenge



— 1 European Union



———— 50 European Healthcare Agencies



———— ~1 x 10<sup>3</sup> Currently Approved Drug Compounds



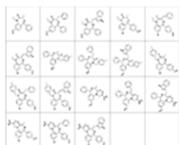
———— ~1 x 10<sup>5</sup> Compounds Used in Commerce and Products



———— ~1 x 10<sup>6</sup> European Researchers



———— ~8 x 10<sup>8</sup> current European Population

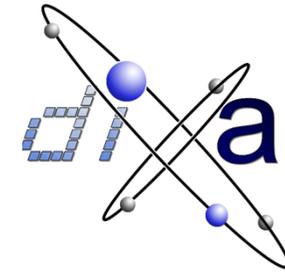


———— ~1 x 10<sup>19</sup> Possible Small Molecule Compounds



## Future expansion of the diXa universe

- **Data explosion in toxicogenomics through advanced 'omics technologies**
  - Gene expression analysis through microarray technologies : several 10s of kB/experiment
  - Proteomics: several 100s of kB/experiment
  - Metabolomics: several MB/experiment
  - Epigenomics: several MB/experiment
  - Next generation sequencing: several TB/experiment
- **Ever expanding domain of chemoinformatics**
  - eChemPortal already captures +25 dBs on chemico/physico /toxicity information
  - ChEMBL covers an excess of 1,300,000 drug-like molecules, over 10,000,000 experimental bioassay results, over 5,000 molecular targets
  - More EU dBs projects' to come (OpenScreen, OpenPhacts, eTOX. ....)
- **Organizing e-Health**
  - Oncology dBs well organized
  - Other disease endpoints ....???



**THANK YOU !!**

[www.diXa-fp7.eu](http://www.diXa-fp7.eu)

[www.icordi.eu/Pages/Home.aspx](http://www.icordi.eu/Pages/Home.aspx)

