



COSMOS Database Public Release Content Summary – December 2013

1. Chemistry

Chemistry data has been collected from various sources. Special emphasis was put on cosmetics related chemicals. The European Union Inventory of Cosmetic Ingredients (CosIng), and the USA Personal Care Products Council (PCPC) inventories have been parsed into COSMOS DB.

The public release contains **81,602 chemical records** and **44,773 unique** (as determined by standard INCHI) **chemical structures**.

The COSMOS consortium is undertaking a concerted chemistry quality control (QC) effort. Approximately 450 randomly selected chemical structures have been subject to QC so far. Based on this effort, a quality assessment of the chemical records in COSMOS is as summarised in Tables 1 through 3 below.

2. Toxicology

The public release of COSMOS DB (December 2013) contains **12,538 toxicological studies for 1,660 compounds**. Two separate datasets are available: US FDA PAFA and oRepeatToxDB. More data(sets) will be added in future updates.

The US FDA PAFA dataset has been generously donated by the Office for Food Additives Safety (OFAS) at the US FDA Center for Food Safety and Applied Nutrition. It contains 12,198 studies **across 27 endpoints including both repeat dose and genetic toxicity data**.

COSMOS consortium harvesting efforts resulted in the assembly of the oRepeatToxDB dataset. Studies from different sources – European Union REACH project, US NTP and others – have been collected. This has resulted in the collection of 340 *in vivo* repeat dose toxicity studies for 228 chemicals. The oRepeatToxDB study records contain the **full plethora of observed toxicological effects** together with the **corresponding sites** at which the effect occurred.

A quality control and assessment over approximately 2% of the records has shown 0.57% erroneous records (e.g., animal counts) and 5.2% missing records (e.g., effect description).



COSMOS Database Public Release Content Summary – December 2013

Table 1. Connection tables QA results

Records QCed	442
Total Corrected CTs / Error rate	43 / 9.7%
Corrected Connectivity / Error rate	16 / 3.6%
Corrected Stereochemistry / Error rate	24 / 5.4%
Corrected Protonation State / Error rate	3 / 0.7%

Table 2. Chemical names QA results

Records QCed	442
Total Corrected Names / Error rate	10 / 2.2%

Table 3. Registry numbers QA results

Records QCed	442
Total Corrected Registry Numbers / Error rate	2 / 0.45%