120,000 structures with cancer/HIV screening data, Derwent World Drug Index WDI—60,000 drugs.

- Toxicity: Aquire—5300 EPA structures with aquatic toxicity, TSCA—100,000 EPA substances
- Reactions: InfoChem ChemReact/ChemSynth—390,000 reactions with 470,000 structures, InfoChem SpresiReact—2.5 million reactions and 1.8 million structures

Software and applications from Daylight include the following (98):

- Numerous toolkits: SMILES, Depict, SMARTS, Fingerprint, Monomer, Thor, Merlin, X-Widgets, Program objects, Remote Access, and Reaction Toolkits (see Glossary)
- Daylight chemistry cartridge for Oracle: DayCart (see Glossary)
- Thor database manager—to build and manage thesaurus-oriented databases
- Merlin searching of structures and data
- Clustering package, with Jarvis-Patrick type cluster analysis
- Rubicon, a program for building 3D models using a distance geometry approach
- PCModels for Log*P* and other physical property calculations
- CombiChem Package to manage high-throughput synthesis
- Reaction Package
- DayCGI—a web development toolkit
- A set of Java tools for chemical information management

Derwent Information. A division of Thomson Scientific, Inc., Derwent is the leading supplier of value-added patent information. The Derwent databases, which are maintained online, include the following:

- Derwent World Patents Index—references to patents, including chemical structure and use patents
- Patents Citations Index—bibliographic and citation data, the Innovations index combined entries from WPI and PCI
- Derwent Selection database—customized subsets of the WPI

The databases are available through several hosting services, including STN, Dialog, and Questel Orbit. User guides for the PCI chemical indexing are available online at Derwent (99). Chemical patents can also be searched using the Merged Markush Service, MicroPatent, and for Japanese patents, the Japanese Patent and Trademark Documents (ISTA) among others (100).

The Gmelin Database. The most comprehensive database of structures, properties, and citations in inorganic and organometallic chemistry is the Gmelin database, based on the Gmelin Handbook of Inorganic and Organometallic Chemistry dating back to 1772. This database includes 1.4 million compounds including coordination compounds, alloys, solid solutions, glasses and ceramics, polymers, and minerals. As such, it is less valuable to drug discovery. The current Gmelin database is owned by the Gesellschaft Deutscher Chemiker and is licensed to MDL GmbH.

MDL Information Systems, Inc. Owned by Elsevier Science Publishing, MDL is a longtime provider of in-house databases and software. Databases include the following:

- Available Chemicals Directory ACD—300,000 structures—reagents and general chemicals, with supplier information
- Bioactivity databases—AIDS database—43,000 structures and data from the National Cancer Institute, Comprehensive Medicinal Chemistry (CMC)—7500 common drug structures, MDL Drug Data Report (MDDR)—120,000 patented drug structures
- Reactions—ChemInform—850,000 reactions and 1.2 million structures, Theilheimer/Chiras/Metalysis—171,000 reactions and 223,000 structures
- Metabolism—Metabolite—53,000 transformations—34,000 structures
- Toxicity—EPA RTECS-based—150,000 structures
- Material safety—OHS Material Safety Data Sheets