



Disposal of nuclear waste in host rock formations featuring high-saline solutions – Implementation of a thermodynamic reference database (THEREDA)



H.C. Moog^{a,*}, F. Bok^b, C.M. Marquardt^c, V. Brendler^b

^a Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) mbH, Theodor-Heuss-Straße 4, 38122 Braunschweig, Germany

^b Helmholtz-Zentrum Dresden – Rossendorf, Institute of Resource Ecology, Bautzner Landstraße 400, 01328 Dresden, Germany

^c Karlsruhe Institute of Technology, Institute for Nuclear Waste Disposal, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

ARTICLE INFO

Article history:

Available online 3 January 2015

ABSTRACT

Research on the solubility of hazardous substances in saturated salt solutions is an ongoing task in Germany. Several institutions deliver contributions in line with their respective expertise. Scientific studies ultimately yield thermodynamic data which are used for thermodynamic equilibrium modelling. In order to join forces and render thermodynamic equilibrium calculations comparable it was decided to setup a common thermodynamic reference database (THEREDA) from which ready-to-use parameter files for commonly used geochemical codes should be created.

It is the objective of this paper to explain how THEREDA is designed from a data management point-of-view, both conceptionally and technically. Data tables and mutual dependencies are described that allow managing administration of data for aqueous solution, solids, solid solutions, and surfaces. Moreover, quality assurance, traceability, consistency, and efficient, long-term maintenance are major topics shaping the database structure. Finally, robust and flexible human interfaces (to editors as well as end-users) are implemented. This paper is not aimed at giving an account of the model definitions, system selections, evaluation schemes, and thermodynamic data themselves stored in THEREDA, which represent the actual scientific work done by many more scientists within the project. However, this methodological guide to THEREDA has its own merits as it helps to bring thermodynamic data to work. Its specific implementation may serve as a useful example for similar projects going far beyond waste disposal.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Safe long term disposal of radioactive waste has been an ongoing task for the last decades and will remain so for decades to come. It is internationally acknowledged that along with site identification and safety analysis the public should be involved and that the procedures for the decision making are reproducible and transparent at all stages.

Part of safety analysis for any candidate site is a reliable estimate about the mobility of toxic waste constituents in the event of water intrusion into direct proximity of the waste containments. Such an estimate is usually performed using thermodynamic equilibrium calculations, often referred to as “geochemical calculations”. Any geochemical calculation rests on four foundations:

- an assumption on the initial composition of the system under consideration;

- an assumption as to the relevant aqueous (or gaseous) species and solid phases;
- a thermodynamic database which contains thermodynamic data for all relevant aqueous (and gaseous) species and solid phases, and
- a suitable code, which upon input of initial system composition and database calculates its composition at thermodynamic equilibrium.

Due to the complexity of systems conceivable upon the contact of aqueous solution with nuclear waste containments (and later on: with the waste matrix itself), the demands on respective thermodynamic databases are challenging. The data must be internally consistent, describe a realistic chemical system, cover wide areas of environmental parameters such pH, redox state, concentrations, temperature or pressure to name just a few, and the datasets should cover all relevant elements and reactions. This implies the use of (defendable) approximations for all those data which are not available at all or exhibit large scatter or uncertainty – the typical tension between rigorousness and comprehensiveness of data.

* Corresponding author. Tel.: +49 531 8012 224; fax: +49 531 8012 200.

E-mail address: helge.moog@grs.de (H.C. Moog).

Research on such thermodynamic data cannot be, and in fact is not, done by a single institution. This led to the situation that several research institutions in Germany maintained their own databases. Missing data were derived from different sources or estimated with different methods. As a result different databases were in use with heterogeneous approaches as to ion-ion interactions, speciation, quality standards and estimation procedures, and stored in different domestic formats which rendered mutual exchange of data and maintenance increasingly difficult.

However, in the event of a final disposition about the mobility of radionuclides in aqueous solutions, differing assessments among several institutions would have a disturbing impact on the overall process of safety analysis and cast doubts on its credibility.

This situation has also been recognized in other countries. As a consequence, efforts to create a thermodynamic database were undertaken i.e.: in the United States of America (“Yucca Mountain Project” (Jove-Colon et al., 2007), development terminated), Switzerland (“Nagra/PSI Chemical Thermodynamic Data Base” (Hummel et al., 2002), France (“ThermoChimie”: <http://www.thermochimie-tdb.com> (Duro et al., 2012; Giffaut et al., 2014) with descendants also used for the Swedish (Duro et al., 2006) and Finnish (Grivé et al., 2008) nuclear waste management programs), Belgium (“MOLDATA” Wang et al., 2010), Japan (“JAEA Thermodynamic DataBase”: <http://migrationdb.jaea.go.jp> (Yoshida and Shibata, 2004), (Kitamura et al., 2012), and the United Kingdom (“HATCHES Thermodynamic Database”: <http://www.hatches-database.com> (Bond et al., 1997), maintenance to be stopped soon in parallel to a set-up of a new database). With the exception of the USA, in none of these database projects high-saline solutions were of special interest.

Therefore, in 2002 plans evolved to develop a common, unified, and obligatory Thermodynamic Reference Database (hereafter referred to as “THEREDA”) for use in safety analysis in Germany with regard to the disposal of heat-generating radioactive waste. It was further decided to focus work on data suitable for the calculation of the solubility of relevant solid phases in high-saline solutions, as they may be encountered in German salt rock and several clay formations. For the database five principal objectives were agreed:

- *Transparency*: any datum should be reproducible in terms of its calculations and traceable to its origin, preferably but not restricted to a peer-reviewed publication.
- *Consistency*: data compiled from different sources should be maintained in an internally consistent way.
- *Comprehensiveness*: all data necessary for performance assessment should be compiled. It follows from this objective that even data with a low quality or even estimated data are to be included if they are necessary.
- *Uncertainties*: if possible, the uncertainty of all entered or internally calculated data should be captured.
- *Data assessment*: along with the capture of data these should be assessed with regard to their quality, their origin, and the method by which they were determined. This also includes the capture of a traceable source from which the data were adopted or derived.

Five institutions joined forces to accomplish this task, any of which represents particular expertise as a benefit for the project:

- Surface Processes Division of the Institute of Resource Ecology at the Helmholtz-Zentrum Dresden-Rossendorf (HZDR-IRE).
- Institute for Nuclear Waste Disposal at the Karlsruhe Institute of Technology (KIT-INE).
- Department for Salt Chemistry of the Institute of Inorganic Chemistry at the Technical University of Freiberg (TUBAF).

- AF Consult, Department of Groundwater Protection and Waste Disposal/Switzerland (AFC).
- Department for Process Analysis of the Final Repository Safety Research Division at the Gesellschaft für Anlagen und Reaktorsicherheit (GRS).

This paper focuses on the technical implementation of THEREDA. The scientific details underlying its contents (thermodynamic data, interaction coefficients, speciation studies, etc.) are published separately by the participating institutions.

2. Long-term considerations

Research for the ultimate disposal of radioactive waste has been an ongoing task for the last couple of decades and will remain so for some decades to come. Therefore, right at the beginning of the THEREDA project decisions were made to ensure, that the principal goal – to safely store an internally consistent set of thermodynamic data and to provide for ready-to-use parameter files – is achieved over a long time scale.

As to the software used it was decided to use open source codes wherever possible; in fact at the time being, THEREDA is entirely build up using open source programs. This way it is made sure, that the database remains functional independent from decisions made by providers of proprietary programs. At the same time it remains possible to adapt used programs to particular necessities, as they may arise in the future.

For the principal way to save thermodynamic data it was decided to use a databank system. A digital relational databank allows using customary mechanisms like foreign keys to ensure internal consistency, enforce certain standards for quality and documentation, and at the same time to allow for an efficient access to a growing amount of data. Basically, foreign keys provide lists, from which items have to be chosen for subsequent entries in other tables. Thus, editors to THEREDA cannot enter information arbitrarily. Foreign keys also require from the editor to comply with a certain workflow, e.g. along the process of declaring a new phase constituent, defining its elemental composition, defining its formation reaction (if appropriate), and assigning values for thermodynamic data.

As to the internal data structure it was essential to compromise on data access efficiency, guarantee for internal consistency, and a degree of data abstraction as low as possible. Ideally, any user with a basic background in chemical thermodynamics should be able to comprehend what is written in each databank table.

To ensure that the principal result of the whole project, the internally consistent database, is preserved for a long time, regardless of ongoing funding, operating systems and databank programs in use, it is essential to have a system which can be archived in plain text. Actually, as will be shown below, THEREDA is archived as script which can be run on other computers whereby the databank is fully restored. To understand the internal structure, an extensive documentation is provided and continuously updated.

The data model was designed in a way to render it adaptable to future needs, even though these may not be fully comprehensible at present. This requirement resulted in the storage of data for both “families” of geochemical codes, Gibbs Energy Minimizers (such as GEMS Wagner et al., 2012; Kulik et al., 2013 or ChemApp Eriksson et al., 1995, 1997) and Law-of-Mass-Action programs (like EQ3/6 Wolery, 1992, PHREEQC (Parkhurst and Appelo, 2013) or Geochemist’s Workbench® Bethke, 2008). Editors to THEREDA may choose which kind of independent thermodynamic data they wish to enter for a given phase constituent (formation- or reaction related). From this dependent thermodynamic data (e.g. formation related) are internally calculated and stored. Editors must select

specific patterns of dependent and independent data from lists. THEREDA is not constrained to any number of (mixed) phases. Any phase can be assigned one or more models for non-ideal mixing effects. There are also no restrictions as to which p,T -functions are internally used, although internally calculations would have to be adapted of course.

3. Basic principles

3.1. Definition

The term “Database” is often used in an unclear manner. For many this term actually means a plain-text ASCII-file which contains thermodynamic data in a code-specific format. Such files can easily be shared, modified with usual text editors and subsequently forwarded to colleagues. If such a file is neither renamed nor documented internally with regard to modifications, it becomes increasingly difficult to judge which version one is working with. Frequently, such files are poorly documented and one can never be sure whether after the addition or modification of data internal consistency is maintained. For such code-specific files we use the name “parameter file”.

“Database” is also used for compilations of selected, recommended data in written form, preferably after a thorough and critical assessment. Two examples are the databases from OECD/NEA (<http://www.oecd-nea.org/dbtdb>) (Mompeán and Wanner, 2003) and PSI/Nagra (Hummel et al., 2002; Thoenen et al., 2004). For both databases code specific parameter files exist, but the functionality to create new parameter files from the modified core-data is beyond the range of normal users.

In view of the practical requirements necessary to create internally consistent and quality assured parameter files we propose a different definition which reflects the way we were approaching our task: A “database” is a technical framework which ensures that dependent and independent data are stored in a manner which ensures that internal consistency is maintained when independent data are modified.

For a definition of dependent and independent data please see Section 6.3.

3.2. Scope

THEREDA in conjunction with a suitable editor's front-end program, hereafter referred to as “DB-Control”, aims at administrating, processing and extracting data intended for usage in programs for the calculation of thermodynamic equilibria in aqueous solutions at temperatures which are of potential interest for hydro-geochemical systems in general and solutions containing hazardous contaminants like radionuclides or heavy metals in particular. As such, THEREDA is not designed to hold experimental data like ThermoML (Frenkel et al., 2006), neither data concerning other liquids than aqueous solution, like melts or other substances which are stable under conditions beyond those where aqueous solutions may exist. Moreover, as for potential applications of THEREDA emphasis is laid on the correct calculation of experimentally determined solubilities, data are preferred which allow the calculational reproduction of such experiments. THEREDA follows the principle of parsimony in the sense that only those solid phases and aqueous phase constituents are adopted which are essential to reproduce solubility data. Reversely, this means that not all aqueous species for whose existence clear evidence exist are included.

The contents of THEREDA are aimed primarily toward applications in the field of nuclear waste disposal in host rock types featuring saline solutions, for which the calculation of activity coefficients is beyond non-ion-specific approaches such as the

Debye-Hueckel model. Thus, at the present state, elements considered in data acquisition are

- Na, K, Mg, Ca, Cl, S, C (system of oceanic salts).
- Al, Si (rock forming minerals, hydrated cement phases).
- Th, Pa, U, Np, Pu, Am, Cm (actinides)
- Cs, Rb, Sr, Ba, Ra, Tc, Sm, Nd (other fission and activation products).

THEREDA does not strive to cover as many solid phases or aqueous species as are described in the literature. Only solid phases are selected which prove to be essential for solubility calculations and which are positively identified, either under natural conditions or in the lab.

For the calculation of aqueous solute activity coefficients the Pitzer model is used in THEREDA (Pitzer, 1991). Thermodynamic data are selected or re-evaluated in order to be consistent with pertinent Pitzer interaction coefficients. These data are usually valid at 298.15 K and 1.01325 bar only. However, for the system of oceanic salts, the validity extends to temperatures up to 473.15 K, depending on the subsystem under consideration.

3.3. Task

THEREDA is not only intended for the long-term administration of thermodynamic data. It is also intended to further communication and collaboration among scientists striving to develop a common, internally consistent database. As this is already difficult enough, THEREDA should facilitate trivial conversion of thermodynamic data or the cumbersome, often prone to errors export for geochemical codes. Data sets can also be “disabled” and are thus not exported. Additionally, the working version of THEREDA is never identical to the version, external users see in the web. Thus, THEREDA is not only used to enter well elaborated, consolidated data, but also to enter tentative values, which after conversion and export for a geochemical code can be tested. Or to put it in other words: THEREDA is also a web-oriented tool for database development.

3.4. NEA data

Experts authorized by the Nuclear Energy Agency (NEA), an institution of the Organisation for Economic Co-operation and Development (OECD), were setting up an internationally recognized reference database for selected actinides and fission products which is periodically – approximately every ten years – extended and updated. Whenever possible, these data were adopted and consistency with NEA-TDB was maintained. Generally, the thermodynamic data at standard conditions were derived from experimental equilibrium constants determined in dilute to concentrated saline systems by using the SIT or the Pitzer equations. The extrapolation of the same experimental data to ionic strength of zero with both approaches often results in slightly different equilibrium constants at infinite dilution. However, the difference between both equilibrium constants is usually smaller than the experimental uncertainty. Whenever both constants are available for the same system THEREDA selects only the equilibrium constant derived with the Pitzer approach, together with the corresponding set of binary and ternary Pitzer parameters. We have not performed an adaption of each approach and we accept that recalculation to real systems with the Pitzer equations yields in a corresponding error. In exceptional cases, especially for high charged ions, a more pragmatic approach has been used. For those cases/systems where data at elevated ionic strengths were also available, these were considered to derive thermodynamic and activity models according with the Pitzer formalism. In some cases, this led to the definition of new species which are only form-

ing at very high ionic strengths and thus are beyond the applicability of SIT (e.g. formation of ternary Ca–An(IV)–OH species). A detailed description on how each individual case is handled would go beyond the scope of this paper, but it can be found in the THEREDA final report (Altmajer et al., 2014) which can be downloaded from the website.

3.5. Technical implementation

THEREDA at its core is a so called relational databank. Its content is structured in tables (“relations”), consisting of columns (“attributes”) and rows (“tuples”). Designing the structure for THEREDA efforts were undertaken to ensure internal consistency on the databank level by enforcing referential integrity between single attributes or even a group of attributes. In practical terms this means for example that many entries are formalized, rendering future queries and data export easier. Another consequence is, generally spoken; that a specific piece of information can only be entered after another one has been entered elsewhere. Two fields of information being linked in this way are sometimes referred to as “parent” (as the information to which is referred) and “child” (as the information which refers to its parent).

However, requirements necessary to render a set of data strictly consistent in terms of thermodynamics are much more complicated than may be realized exclusively by means of databank logic. Apart from the actual databank, other programs are used, to administrate, access, view, and export contents of the databank. Therefore, THEREDA may be rather thought of as a suite of programs for the storage and administration of thermodynamic and related data. This suite entails:

- The databank, as represented by the data model and a system of checks and foreign keys ensuring referential integrity.
- Scripts which perform internal calculations.
- Scripts to offer views according to particular criteria and detect data deficiencies.
- A JAVA-based program to export selected data into a generic JSON-format.
- PHP-scripts to convert JSON-files into code specific parameter files.
- A website which along with a content management system facilitates project-internal communication as well as providing an interface to interested users.
- A web based frontend (DB-Control) to enter and administrate the stored thermodynamic data.

For the reasons given, part of the logic necessary to produce consistent and meaningful datasets are located in the databank itself, while other parts must be placed within the code of DB-Control or as triggers attached to relations in the databank. These parts comprise issuing of SQL-statements to compute constraints for entered data which otherwise cannot be implemented as part of the databank structure itself (without undue complexity).

As format for the databank we chose PostgreSQL. It is a powerful, open source relational database system, which originates from a development at the University of California at Berkeley in the 1970s. Behind PostgreSQL are more than 20 years of active development and a proven architecture that has earned it a strong reputation for reliability, data integrity, and correctness. It runs on all major operating systems, including Linux, UNIX (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, Tru64), and Windows.

To avoid an undue length of this article, the complete structure of the databank cannot be given in full detail here. For anybody interested we refer to the Technical Documentation of the THEREDA Relational Database Management System (RDBMS), available for download from the THEREDA website.

3.6. General information

Generally, most data sets in THEREDA are provided with dedicated fields for additional descriptions intended for the user, or remarks for internal use. Most data sets have an Owner (the originator of the data set) and a time stamp, which allows tracking down changes, possibly done during work with THEREDA.

3.7. Permissions

THEREDA can only be operated by persons having the status “editor”. An editor has the permission to create new data sets. Although all data sets can be read by everybody in the editors-team, the system prevents the change of a data set by anybody else than the editor. Thus, THEREDA relies on shared responsibilities: editors “own” data sets. They are responsible that “their” data were entered according to mutually agreed guidelines, were correctly adopted from literature and are well documented.

3.8. Release politics

At the time being, data releases are aligned to specific subsystems, for which test calculations were performed with four different geochemical speciation codes. Data releases so far covered the systems

1. Na⁺, K⁺, Mg²⁺, Ca²⁺–Cl[–], SO₄^{2–}–H₂O(l) (polytherm).
2. Na⁺, Mg²⁺, Ca²⁺–Cl[–]–Am³⁺, Nd³⁺, Cm³⁺–H₂O(l).
3. Na⁺, Mg²⁺, Ca²⁺–Cl[–], SO₄^{2–}–CO₃^{2–}/CO₂(g)–H₂O(l).
4. Na⁺–Cl[–]–NpO₂²⁺–H₂O(l).
5. Na⁺, K⁺, Cs⁺, Mg²⁺, Ca²⁺–Cl[–], SO₄^{2–}–CO₃^{2–}/CO₂(g)–H₂O(l).
6. Cement phases – Si(OH)₄, Al(OH)₄[–]–Na⁺, Mg²⁺, Ca²⁺, K⁺–Cl[–], SO₄^{2–}, CO₃^{2–}–H₂O(l).
7. Na⁺, K⁺, Ca²⁺–Cl[–]–Th⁴⁺, Np⁴⁺, Pu⁴⁺–CO₃^{2–}–H₂O(l).
8. Na⁺, Mg²⁺, Ca²⁺–Cl[–], SO₄^{2–}, CO₃^{2–}–Am³⁺, Cm³⁺–H₂O(l).
9. Na⁺, K⁺, Mg²⁺, Ca²⁺–Cl[–], SO₄^{2–}, CO₃^{2–}–Si(OH)₄–U⁴⁺/UO₂²⁺–H₂O(l).

Some more data releases are being prepared at the time of writing this article, each one dedicated to another, rather small subsystem. Thus, the amount of data released so far is much less than the data altogether captured in THEREDA. There has been no “bulk release” of the entire database, and therefore THEREDA does not feature yet a “version number”.

Further development of the database will proceed, by releasing (or adding) further subsystems and by examining possible interactions between existing releases. The latter procedure might lead to the addition of new data. E.g., a data release for the system Na⁺, K⁺–Cl[–], SO₄^{2–}, HPO₄[–]–H₂O(l) (which is due in 2015) could be combined with release number 6, leading to the examination of new solid phases with U(+VI) or U(+IV).

In the long term, data releases from THEREDA will gradually evolve into more complex, and consequently more comprehensive data sets. Necessarily, this will eventually render earlier data releases obsolete and ultimately lead to a smaller number of released parameter files. Once the most relevant chemical systems are covered the release strategy will be revised, probably to biannual releases of the database that are uniquely marked and incorporate all cumulative changes since the previous release. It is only in this stage of development that issuing of version numbers for THEREDA will become meaningful.

Please note, that regardless of future decisions as to the general release strategy of THEREDA, parameter files once released along with the whole database are archived for the entire life span of THEREDA. Thus, older computations can be repeated at any time.

4. Basics

In this section it is explained how the chemical material base is setup to which all data refer. An illustration of the involved tables and their crosslinking is given in the [supplements](#). A list of elements and phases form the root. For any phase additional properties may be entered as appropriate, e.g. mineralogical properties or trivial names. Each phase is composed of phase constituents, for which the elemental composition is defined. This means, that on this level of THEREDA all mixed phases, including the gaseous, the aqueous and all solid solutions are treated alike. Also, phase constituents are defined in terms of a formation reaction, when appropriate.

To ensure uniform usage among all editors of THEREDA basic physical constants are stored as internal reference. This relation is not linked to any other relation in the databank. These values are also used for internal calculations, if necessary. They were adopted from [Hummel et al. \(2005\)](#) and [Fuger et al. \(2008\)](#) and are listed in the [supplement](#) provided with this paper.

Chemical elements are stored in a dedicated relation. It contains i.a. symbols, names, molar masses, standard entropies and heat capacities. For internal calculations it also contains the number of atoms in the elementary compound to which the standard entropy refers. E.g. for the element nitrogen this would be gaseous N₂ with the coefficient 2.

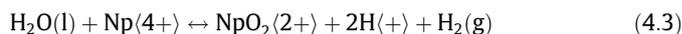
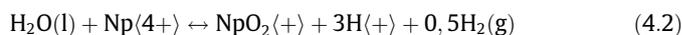
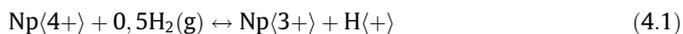
Phase constituents are declared in a dedicated relation “PCon” and assigned to formerly declared phases. The naming of phase constituents is not done arbitrarily but follows documented rules. For a detailed description of the nomenclature that is used in THEREDA, see the THEREDA Technical Paper “Guideline for the creation of identifiers of phase constituents and phases” which is freely available at the projects website <http://www.thereda.de>.

This relation may be understood as a kind of fork as many other relations refer to it. By this relation all kinds of data are associated with a particular phase constituent, and as such form a set of data which are spread over a number of relations in the databank. The elemental composition and charge of any phase constituent is defined in relation “PConComposition”. The molar mass is calculated internally.

All phase constituents are assigned a type, as listed in [Table 4.1](#).

“PrimaryMaster” are not defined by a formation reaction. “SecondaryMaster” in THEREDA are commonly those phase constituents which are formed by reduction or oxidation from a “PrimaryMaster”, and which in turn constitute the educt for further reactions to form products, where the redox sensitive element remains in the same oxidation state. This is made clearer in the following example.

Np (4+) is “PrimaryMaster” for the element Neptunium. Phase constituents Np (3+), NpO₂ (+), and NpO₂ (2+) are formed from Np (4+) and thus are declared as “SecondaryMaster”:



Phase constituents from reactions between “PrimaryMaster” and/or “SecondaryMaster” are declared as “Product” or “MineralsSolids”. E.g. phase constituent NpO₂(CO₃)₃ (5−) is formed from NpO₂ (+) and thus is declared as “Product”:



All constituents of solid phases must be of type “MineralsSolids”. In THEREDA phase constituents of type “Product” or “MineralsSolids” cannot occur as educts in a reaction. Note that all reactions in THEREDA are entered as formation reactions of the phase constituent or solid phase in question. All reaction related thermodynamic data stored in THEREDA refer to these reactions.

Formation reactions stored in THEREDA do not necessarily reflect real reaction mechanisms. They create a formal mathematical framework which governs how dependent data are internally calculated from independent ones. Upon export for other geochemical codes these reactions are transformed as necessary, e.g. recast in a reaction using “aqueous electrons” e(−) or dissolved O₂(0).

The described classification of phase constituents into types has two major consequences for THEREDA. First, not all types of data are suitable for any type of phase constituents. E.g., for a “PrimaryMaster” no reaction related data types such as log *K* exist by definition, and must therefore not be entered, and in fact cannot thanks to internal check routines. Furthermore, in organizing phase constituents into a hierarchy, internal calculations are only done for those phase constituents whose data are affected by a change in data for another one. Closer details are explained in [Section 6](#).

Phase constituents are sorted into phases, using a dedicated relation called “Phase_x_PCon”. Thus, an arbitrary number of mixed phases can be stored in THEREDA. Any phase for which more than one end member is registered is cleared by the system for the assignment of non-ideal mixing models. On this level it does not matter which aggregation state a phase has. Note, that the same formalism is maintained for pure stoichiometric phases with only one end member: this means that phase constituent “NaCl(cr)” is assigned to phase “Halite”. Another consequence of this structure is that a phase constituent can belong to more than one mixed phase.

As a tribute to the usual codes for thermodynamic equilibrium calculation, “SecondaryMaster” (“redox species” in other terms) need to be stored in conjunction with their respective central elements and their oxidation states. The situation is further complicated by the fact, that some solid phases may contain more than one redox-variable element (such as FeS₂(cr)), and – even more complicated – may even contain the same element in two different oxidation states (such as Fe₃O₄(cr)). It is for these cases that the assignment of redox-sensitive elements and their respective oxidation states is left to another separate relation called “OxidationState”.

Even after capturing all necessary data it is still up to the final judgment of the editor whether a phase constituent or a phase should be included in a data release or not. The field “Enabled” in relation “PCon” serves as final gate. This function allows extending the database continuously without being forced to release those parts which are still under development. Note that an “enabled” phase constituent can still be commented out in a release due to equilibrium constraints. This is necessary to transmit the information to the user, that a particular species or solid phase has not been “forgotten” but that its usage is discouraged for reasons given in the description field.

Table 4.1

Relation “PConType”: permitted types of phase constituents and their description.

Symbol	Description
PrimaryMaster	Only formation data, not defined by any reaction
SecondaryMaster	Defined by reaction of primary master phase constituents
Product	Defined by reaction of primary and/or secondary master phase constituents
MineralsSolids	For phase constituents of solid phases; defined by reaction of primary and/or secondary master phase constituents
SurfacePrimary	PrimaryMaster for surface complexation reactions
SurfaceProduct	Product for surface complexation reactions

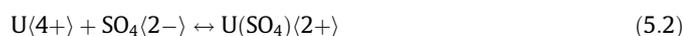
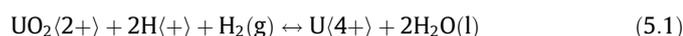
For documentary reasons an attribute called “OriginalReactionEquation” is used to store the originally published reaction equation if it differs from the one imposed by THEREDA due to the allowed (primary and secondary) master species.

5. Reactions

An illustration of the involved tables and their crosslinking is given in the [supplements](#). Phase constituents in THEREDA are grouped into four types, as declared in relation “PConType”. The only phase constituents not defined by reaction are of type “PrimaryMaster”. This means, in turn, that all other phase constituents, which are not of type “PrimaryMaster”, MUST be defined by a reaction. This reaction, by convention among the editors, is always the forming reaction of the phase constituent to be defined. Note that in this scheme also surface complexation reactions are covered.

Reactions entered in THEREDA are exported into parameter files dedicated for law-of-mass-action programs. Therefore, it is essential that for any phase constituent which is not “PrimaryMaster” only one forming reaction is defined.

To clarify the system how reactions are entered in THEREDA, a table provided in the [supplements](#) shows how the forming reactions of $U(4+)$ and $U(SO_4)(2+)$ are defined.



Note that redox reactions in THEREDA are defined uniformly in terms of $H_2(g)$. However, upon export these redox reactions are transformed for other redox species (such as “ $O_2(aq)$ ”) if necessary.

A third example shows the formation of Galena (PbS) according to the following reaction:



Note that the principal product of reaction as declared in field “PCon_Product” is also entered into field “PCon_Reactant”. Generally, the described system to store reactions facilitates the calculation of reaction related standard thermodynamic data efficiently.

Generally it is assumed, that a reversible equilibrium exists for any reaction-defined phase constituent in THEREDA – kinetic constraints are not dealt with so far. However, solid phases exist, for which proper thermodynamic data are available, but which as to the best knowledge of experimenters are not likely to be in equilibrium with an aqueous solution. Formation/precipitation or dissociation/dissolution can be subject to kinetic control or inhibited altogether. Having such phases in calculations may lead to erroneous results. Therefore, such phase constituents are marked accordingly in THEREDA (see [Table 5.1](#)). The same applies to some redox reactions, which are appraised as usually being not in equilibrium. The following table gives a list of presently permitted entries.

At the time being 22 phase constituents are marked as “No equilibrium”, e.g. solids like $Tc_2O_7(cr)$. Such phases may even turn up in released parameter files but are commented out with an appropriate note for the user.

Table 5.1
Relation “Equilibrium_Constraint”: permitted entries and their description.

Symbol	Description
Formation	Formation/precipitation is at equilibrium only
Dissociation	Dissociation/dissolution is at equilibrium only
No equilibrium	Neither precipitation nor dissolution is at equilibrium
Complete equilibrium	Both, dissociation/dissolution and formation/precipitation is always at equilibrium

6. Data

An illustration of the involved tables and their crosslinking is given in the [supplements](#).

6.1. General structure

At present all thermodynamic data released are designed to be consistent within the Pitzer framework for high saline solutions (for explanations see [Section 3.4](#)).

Data valid at 298.15 K and a pressure of 1.01325 bar are stored in relation “data_standard_pitzer”, while p,T -functions are stored in relation “data_variable_pitzer”. As in Germany no final decision has been met up to date as to the host rock intended for the disposal of nuclear waste, THEREDA contains two more relations to hold SIT- and Extended-Debye-Hueckel consistent data. These data sets, though in great parts identical to those in the Pitzer part, are not further developed at present. They can be regarded as different strings of data: internal calculations are performed only within each string. They are held together by relation “PCon”. Thus, any phase constituent can hold different sets of thermodynamic data, each one consistent with a particular model for the calculation of activity coefficients. This has consequences for equilibrium constants whose extrapolation to the standard state can lead to different values, depending on the model used (Pitzer, SIT, or Debye-Hueckel). In any data table, the databank enforces uniqueness for any combination of phase constituent and data type. This way internal calculations and export of data to parameter files can proceed unambiguously.

It is important to note, that the presence of data for any given model (i.e. Pitzer, SIT or Debye-Hueckel) effectively determines the speciation model for the aqueous solution. For example, thermodynamic data for some ion pairs may be present in the table for Debye-Hueckel data, but not in the other data tables. Consequently, these species would be exported only into parameter files employing the Debye-Hueckel model, but not in others.

In relation “data_variable_pitzer” up to six coefficients for “DataType” can be saved in subsequent fields, the meanings of which are defined by the setting of “TPFunc”, i.e. a symbol for the p,T -function in mind.

Field “TState” provides a list for the selection of temperature and pressure state identifiers of data type. Basically there are only two of them: any data type is EITHER valid at 298.15 K and 0.1 MPa only (“standard” state or “S”) OR refers to a function for variable temperature and pressure conditions (“variable” state or “V”). It is usually set automatically by the system.

All thermodynamic data entered in THEREDA are assigned a “DataType”. A table in the [supplements](#) gives a list of currently permitted data types. All data types containing a “298” in their designator refer to the standard temperature $T = 298.15$ K.

Relation “TPFunc” provides a list of permitted temperature and pressure functions for thermodynamic parameters or interaction coefficients. A table in the [supplements](#) gives a list of presently permitted p,T -functions.

These equations reflect functions which are used by the supported geochemical equilibrium codes or are otherwise necessary for internal calculations. This list can be prolonged if necessary. Though the name “TPFunc” might suggest otherwise, at present no pressure functions are in use in THEREDA. However, THEREDA is prepared to hold standard molar (for solids) or partial standard molal volumes (for aqueous species) to account for pressure corrections for chemical equilibria.

Whenever a temperature function is entered in THEREDA, it is mandatory to define the range of temperature for which it is valid. This information is specific for any phase constituent for which it is

entered and is transmitted to the user upon export into code-specific parameter files.

Units for all kinds of data cannot be selected by the editor. Usually all units are SI-based or were selected on the mutual acceptance of the THEREDA administration board. However, each data type is assigned a meaningful acronym for its unit, which is displayed to the editor by DB-Control.

Generally, all decimal numerical values are inserted with “.” instead of “,”. In SQL-statements the latter is interpreted as separator between attributes.

6.2. Classification and categorization of data

Data classification in THEREDA consists of a numerical element (the data class) and the data category. The number is a measure of the overall credibility of the datum. Data category specifies how the datum was determined experimentally or derived calculationally. It must not be confused with the calculation mode described below. While the numerical part refers to more general information about the datum, data category identifies whether it refers to formation or reaction data. Table 6.1 gives a list of presently permitted categories.

Background for the categorization of data is our experience that e.g. a solubility constant for a solid phase can sometimes be traced back to an experimental determination of standard enthalpy, standard entropy, and heat capacity with purely thermochemical methods. This value, however accurate the measurement might have been performed, can turn out to be completely unusable for the calculation of solubility in aqueous solution. However, the purpose of THEREDA is to compile data for the modelling of aqueous solution chemistry. Therefore, data based on equilibrium data in aqueous solution are preferred normally.

However, as inevitably not all data needed for solubility equilibria in aqueous solution will be available, contributors to THEREDA may decide to accept a datum, which was derived thermochemically. But in any case the thermochemical nature of its origin will be saved in THEREDA for both documentation and to issue appropriate warnings to the user.

Entries for data class and data category are selected from lists. Table 6.2 gives a list of presently permitted combinations of data class and –category.

This system reflects the different steps in which data are entered. “DataClass”=6 is valid as soon as a data set has been automatically generated upon definition of a new phase constituent. It usually means the editor has not yet touched the data set at all. In a second step data are entered. The assessment about the quality and the source (see below) can be done along with the entry or later on. Categories “F” and “R” refer to values derived from formation or reaction data, respectively. In some cases this categorization, however, is not applicable (“NA”).

The general meaning of the numbers reflecting the data quality is given in Table 6.3.

Data are also classified as to their origin (“DataSource”). A system of numbers specifies the kind of reference a datum is taken from. Table 6.4 gives a list of presently permitted data sources.

Table 6.1
Relation “Category”: permitted categories and their description.

Symbol	Description
F	Formation data
R	Reaction data
FR	Based on formation- and reaction-type data
S	Surface-related data
NA	Not applicable

Looking at the items in the tables for data class, data quality, and data source, it becomes obvious that THEREDA is actually prepared to accept data with a low credibility. Such data are adopted if they are essential for certain geochemical calculations and no better data are available. Accepting highly reliable data only would bring the user to amend the exported parameter file on his own initiative. While this of course remains possible at any time, we felt it preferable to propose such data straightaway. The important point is, however, that the information about poor data quality is passed on to the user in the exported parameter files!

All data in THEREDA refer to a sourceable reference. References are stored in a dedicated substructure, explained in detail in the [supplements](#).

6.3. Internal calculations

Basically, all kind of data in THEREDA may be either “entered” (“independent”) or “calculated” (“dependent”). In the former case, the datum is considered fixed and not subject to internal calculations. Note that an “entered” datum may still be a calculated one, but this calculation was done “outside” the databank. In the latter case, the datum is derived from other data in THEREDA by internal calculations.

However, the exact way HOW a datum is to be derived from other data becomes unambiguous only as soon as “CalcMode” is combined with a “DataType”. Not all calculation modes make sense for all data types. And not for all possible combinations of data type and calculation mode (even though they may be feasible) an algorithm may be available in THEREDA. Only particular combinations of “CalcMode” and “DataType” are allowed, which are declared in the n:m relation “DataType_x_CalcMode”. Actually, for some data types (“DRG298”, “DFG298”), it is not permitted at all to enter values directly. A table in the [supplements](#) gives a list of presently permitted calculation modes.

Upon the definition of a new phase constituent, the system creates new standard combinations of entered and internally calculated data types, which is to be selected by the editor.

As THEREDA is an ongoing project, some internal calculations, though meaningful, are not implemented yet. This is transmitted in a dedicated field to the editor.

As to internal calculations of dependent data, THEREDA follows the concept to define a minimum number of necessary, elemental conversion steps between certain data types. By repeating those steps all over the entire databank, eventually all dependent data are calculated. Fig. 6.1 shows how the different conversion steps create links between all data types.

Generally, internal calculations work with the usual thermodynamic relations. E.g., data types in boxes represent those which may be mutually calculated by use of the Gibbs–Helmholtz equation of formation or reaction, whenever two out of three are known. “CF” represents the calculation of reaction data from formation data. Conversely, calculation of formation data from reaction data is marked “CR”.

A closer inspection of Fig. 6.1 reveals that certain conversions are performed in a succession of elemental calculational steps. E.g., “LOGKT” is not directly converted into “LOGK298”, as might have been expected, but first converted into “DRGT”, which subsequently is converted into “DRG298”. It is with “DRG298” only, that calculation of “LOGK298” is available (when it has not been ENTERED).

“CP298”, the standard (partial) molar heat capacity at constant pressure is entered directly into the databank. It should be kept in mind, however, that for the derivation of heat capacities in the cited literature, other standard thermodynamic quantities may have been used. This might especially hold true for the derivation

Table 6.2

Relation “DataClass”: permitted combinations of data quality (column “Symbol”) and data category and their description.

Symbol	Category	Description
–1	F	Internally calculated with CalcMode CF, CGHF or CTPFUNC
–1	R	Internally calculated with CalcMode CR, CGHR or CRLOG
–1	NA	Value internally calculated
0	NA	By definition/convention fixed value
1	F	Value based upon experimental thermochemical data
1	S	Value based upon experimental equilibrium data
1	R	Value based upon experimental equilibrium data in aqueous solution
2	F	Chemical analogue value, based upon thermochemical data
2	S	Chemical analogue value
2	R	Chemical analogue value, based upon experimental equilibrium data in aqueous solution
3	F	Estimated value, based upon founded correlations and models for formation data
3	S	Estimated value, based upon founded correlations and models
3	R	Estimated value, based upon founded correlations and models for reaction data
3	NA	Estimated value, based upon founded correlations and models
4	NA	Origin of value not reported; data class cannot be determined
5	NA	Not consistent with other data in THEREDA
6	NA	DataClass has not yet been assigned

Table 6.3

Relation “DataQuality”: permitted data qualities and their description.

Symbol	Description
–1	Internally calculated
0	By definition/convention fixed value
1	Reliable datum
2	Datum is reliable within the given range of error, but error is relatively high (because of experimental problems, errors in utilized auxiliary data, or uncertainties due to inappropriate analogy-data or methods of estimation)
3	Questionable value (uncertain model for speciation, uncertain auxiliary data), but nevertheless suitable and necessary for the description of experimental data in the system of interest
4	Suitability for modelling or correctness not yet determined
5	Scrutinized and deemed inapplicable for modelling (due to experimental shortcomings or inadequate assumptions in the course of processing experimental data or inadequate estimation procedures)
6	DataQuality not yet entered (to be done)

Table 6.4

Relation “DataSource”: permitted data sources and their description.

Symbol	Description
–1	Internally calculated
0	By definition/convention fixed value
1	Value taken from an international review (e.g. CODATA, NEA-TDB) or from an internationally acknowledged review article
2	Value taken from an institutional review
3	Value is based upon a number of publicly accessible publications (paper, report)
4	Value is based upon a single publicly accessible publication (paper, report)
5	Value is based upon internal sources not available to the public, but available to editors of THEREDA
6	Data source not yet entered (to be done)

of partial molal standard heat capacities of aqueous species (e.g. Bernarducci et al., 1979).

Another special case is the relationship between “DRG298” and “LOGK298v. It is possible, though not advisable, to derive “LOGK298” from “DRG298” (“CR”), which is calculated from “DRH298” and “DRS298” (“CGHR”), which in turn were derived from “DFH298” and “S298” (“CF”). This means, while the final calculational step “CR” may lead to the conclusion that “LOGK298” is derived from reaction data, it is ultimately derived from formation data (“DFH298” and “S298”), which clearly violates one key objective of THEREDA: to unambiguously mark each datum, whether it is based on aqueous equilibrium data or thermochemical data. This

predicament was solved by introducing a new elemental calculational step “CF”, which in this context leads to exactly the same mathematical transaction like “CR”.

Missing in the above figure are interaction parameters (“IP298”, “IPT”), which are either entered or calculated from a p,T -function. Data types “IP298” and “IPT” are generic and can contain all kinds of interaction coefficients for any phase. How these coefficients are internally calculated is governed by the interaction model and the interaction type of the coefficients in question, see Section 7.

Internal calculations are triggered by the change of any data type with “CalcMode” = “Entered”. Basically, scripts written in PL/pgSQL (procedural language/PostgreSQL structured query language) begin to work. PL/pgSQL is supported by the PostgreSQL database management system (DBMS) and offers the possibility to combine SQL-statements with syntax elements which allow for the construction of loops and control structures. Thus, PL/pgSQL-scripts allow issuing queries, to perform arithmetic calculations from them and to update databank tables.

The main script is called with the name of the data table and the present phase constituent as arguments. It is checked whether the data set changed is actually a temperature function or a value valid for 298.15 K only. In the former case, dependent temperature functions are internally calculated first, followed by depending data sets for 298.15 K. Following the hierarchical structure of phase constituents it is determined which other data depend on the one just changed. If the changed datum belongs to a phase constituent of type “Product” or “MineralsSolids”, only data belonging to this phase constituent are subject to change, because – by defini-

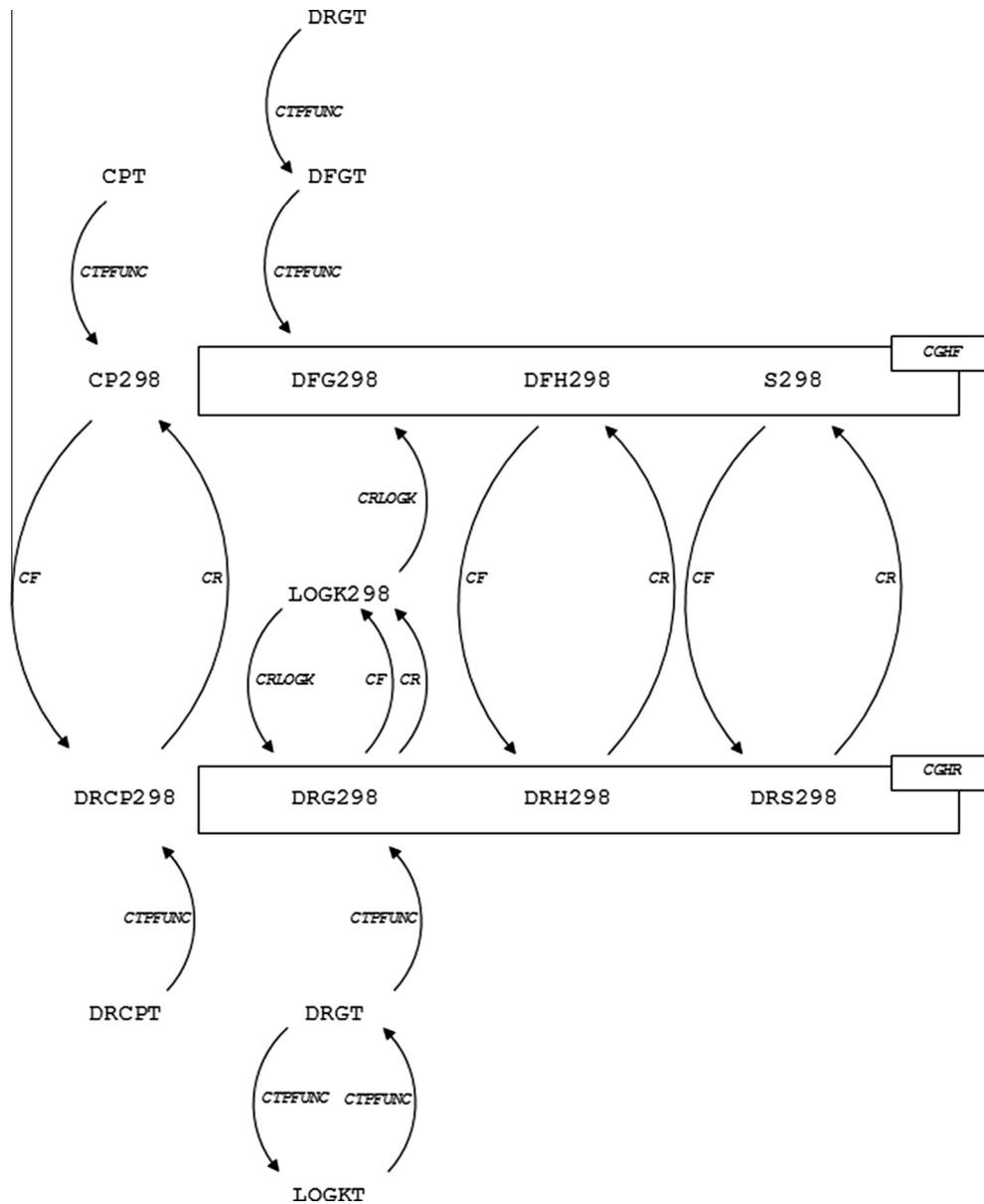


Fig. 6.1. Elemental conversion steps between data types.

tion – no other phase constituents depend from it. If a “Secondary-Master” has been changed, first all data for this phase constituent are re-evaluated; then a list of depending phase constituents is created and the main function loops through the list to calculate new data for all phase constituents on this list. The same applies in case a “PrimaryMaster” is changed (which is, however, almost never the case): after re-evaluation of all data for the “PrimaryMaster”, a list of all depending “SecondaryMasters” and “Product”/“MineralsSolids” is formed and re-evaluated. The list of data types belonging to a single phase constituent is looped through until all dependent data for the respective phase constituent are calculated.

It is important to note that all internal calculations are confined to data belonging to the present interaction solution model. For example, if a datum belonging to the Pitzer database is changed (“data_standard_pitzer” or “data_variable_pitzer”) then internal calculations are performed only for other data in these relations. This way, work on two (or more) different databases could proceed

simultaneously, maintaining internal consistency and referring to a single system of phase constituents, classification and formation reactions. Note, that this system also allows having different speciation models for different aqueous activity models: only phase constituents which are assigned thermodynamic data for a given aqueous activity model (may it be Pitzer, Extended Debye-Hueckel or other models) are exported into the respective parameter files for geochemical speciation codes.

To reduce the complexity and length of code used, each elemental conversion step, depicted as arrows in Fig. 6.1 is represented by an individual function. Any function checks whether the prerequisites for the calculation are met. If this is not the case, a message is dropped for the editor in field “NoteForEditor” (which exists in all data tables) and the respective value (or all values for temperature coefficients, respectively) is set to “NULL”. If the calculation is possible, “NoteForEditor” is set to “NULL”, and the table is updated with the new value(s).

7. Interactions

An illustration of the involved tables and their crosslinking is given in the [supplements](#).

THEREDA uses a very general approach for non-ideal interactions which is applied to all phases, the aqueous, the gaseous, and all solid phases. Administration of interactions within THEREDA begins with the declaration of an interaction model. [Table 7.1](#) gives a list of currently permitted interaction models. “EDH” is present for formal reasons: The extended Debye-Hueckel model does not contain interaction coefficients of any kind.

An arbitrary number of additional interaction models can be added, including ones for non-ideal solid solutions. Note, that the Redlich–Kister–Muggianu–Polynomial has been added to the list, to test the databank functionalities during development. However, at present no real data for non-ideal solid solutions are part of THEREDA, because it was not necessary yet.

Any model may be assigned an arbitrary number of interaction types and must be assigned to at least one mixed phase. This construction works similar to the assignment of phase constituents to phases in n:m relation called “InteractionModel_x_Phase”. For example, the Pitzer model is assigned to the aqueous phase only, and thus the databank would prevent the entry such data for a solid mixed phase. Another example would be the RKMP-model, which could be assigned to a number of solid solutions. It is only after this assignment that a particular phase is cleared by the system for the entry of pertinent data.

Next step is the declaration of interaction types. Each interaction type is assigned to an interaction model, a number of interacting phase constituents, and the number of coefficients, by which it is defined. A table in the [supplements](#) gives a list of currently permitted interaction types.

Interaction types related to the SIT-theory were included to test functionalities of the databank. At present, no activities are undertaken to develop a SIT-consistent database in THEREDA.

Any interaction is defined by a particular combination of phase, interaction type and interacting phase constituents. Such interactions are declared in relation “Interaction”. It is only after these steps that interaction coefficients (being associated with a particular interaction type which is part of a particular interaction model) may be entered.

In analogy to the declaration of data classes, the relation “IPClass” provides a list of identifiers for classes of interaction parameters, which are better suited than those designed for the classification of data. Note that “IPClass” is essentially built up by one attribute only while data classes are characterized by a combination of two attributes (compare Section 6.2) (see [Table 7.2](#)).

[Table 7.2](#) gives a list of presently permitted classes for interaction parameters.

As for thermodynamic data, interaction coefficients are recorded for 298.15 K or as functions of temperature and pressure in relations “interaction_standard” and “interaction_variable”. Up to six coefficients can be stored for any individual interaction. Note, that coefficients for conceivable interaction models stored

Table 7.1
Relation “InteractionModel”: permitted interaction models.

Symbol	Description
Pitzer	Pitzer model
RKMP	Redlich–Kister–Muggianu Polynomial
SIT	Specific Interaction Theory
EDH	Extended Debye-Hueckel

Table 7.2
Relation “IPClass”: permitted classes for interaction parameters and their description.

Symbol	Description
–1	Internally calculated with CalcMode CTPFUNC
0	By definition/convention fixed value
1	Value based upon experimental equilibrium data
2	Chemical analogue value, based upon experimental equilibrium data
3	Estimated value, based upon founded correlations and models
4	Tentative value for unknown interaction coefficients which cannot be estimated
5	Not consistent with other data in THEREDA
6	IPClass not yet entered (to be done)

in THEREDA are stored in these two relations, unlike the situation with standard thermodynamic data, which are stored in different relations, depending on their consistency with a particular aqueous solution model.

In “interaction_variable” up to six temperature (or pressure) coefficients may be entered for each of the six 298 K-coefficient. Again, the meaning of these coefficients is given by the selected *p,T*-function. However, at present only T-functions for interaction coefficients are in use.

Note the similarities with the way standard thermodynamic data are stored. Also for interaction coefficients information as to the quality, the category and the source of the interaction coefficients is captured. Additional information covers the valid range of temperature and pressure, as is done for standard thermodynamic data.

At the time being, no solid solution models are entered in THEREDA. Solid solutions entered without the assignment of interaction coefficients would be treated as ideal solid solutions upon export.

8. Surface complexation

In addition to the precipitation of and incorporation in mineral phases, also the sorption onto mineral surfaces is an important retardation process for radio-/chemotoxic substances on their possible way from a nuclear waste deposit into the biosphere. Surfaces to be considered include the backfill and the surrounding host rock as well as corrosion products of the various barrier components. For the modelling of such sorption processes various Surface Complexation Models (SCM) have been established. These SCM are not consistent to the Pitzer ion-ion interaction model (primarily addressed by THEREDA so far) but to the Extended Debye-Hueckel (EDH) interaction model. However, the database structure has been extended for the processing of surface complexation data of the following SCM's:

- Diffuse double-layer model (DDL) [Dzombak and Morel \(1990\)](#).
- Constant Capacitance model (CC) [Schindler and Gamsjäger \(1972\)](#).
- Non-electrostatic model (NE) [Bradbury and Baeyens \(1997\)](#).
- 1pK-Basic Stern model (1pK-BS) [Westall and Hohl \(1980\)](#).

This implementation is fully consistent with the classification and categorization system as described in Section 6 including internal calculations (see Section 6.3).

So far, no efforts were undertaken within the frame of THEREDA to develop an internally consistent database for low-saline solutions. Thus, SCM data support is “under development” and might be subject to future versions of THEREDA. The specific technical implementation will be described in a later publication.

9. Sets

An illustration of the involved tables and their crosslinking is given in the [supplements](#).

Maintaining “internal consistency” means to obey certain relationships between individual data, going well beyond the usual thermodynamic relationships such as Gibbs–Helmholtz, Hess or Born–Haber. The easiest way to indicate this is to put individual data into a single data set. Examples are the six individual interaction coefficients for a single binary Pitzer interaction only being valid for a certain solubility constant, or the individual terms for p,T -function in “Data_Variable_Pitzer”. In either case, individual data pertaining to a single data set depend on each other and thus must only be changed simultaneously and in a consistent manner.

A more intricate way is given by the THEREDA system of internal calculations, where all data either “entered” or “internally calculated”: as with the former case, an internally calculated value must not be changed manually (and in fact cannot be in THEREDA), because it depends from an entered value elsewhere. Vice versa, upon changing an entered value, all dependent internally calculated values need to be updated (this is in fact done by calculation scripts within the databank).

A third example is represented by the very relation structure within THEREDA, where data consistent with the Pitzer approach are asserted into one individual relation (“Data_Standard_Pitzer” or “Data_Variable_Pitzer”). The same would apply, if THEREDA would be enhanced to cover also an Extended-Debye-Hueckel model. As a consequence, a mineral phase in THEREDA may have as many solubility constants as there are different strings of aqueous phase models present: a solubility constant in “Data_Standard_Pitzer” is consistent with all other data pertaining to the Pitzer world, while a solubility constant for the same mineral phase consistent with Debye-Hueckel data could be numerically different, and yet leading to correctly calculated solubilities when applied in a geochemical code.

In all these cases individual data form what we call “sets”, a term which should be clearly distinguished from the technical term “data sets”, which actually means “tuples” in the data bank world. However, dependencies between data may exist, which cannot be represented in a databank structure without undue effort. For this purpose, THEREDA offers the possibility to declare named “sets” (relation “SetList”), and to assign interactions, phase constituents, phases or data arbitrarily (relation “SetData”). The purpose of this concept is to notify the editor in the event when individual data are about to be modified which depend from other data or from which other data depend, respectively.

A practical example for such a set is the solubility constant for a solid phase, which had been developed along with Pitzer parameters, describing interactions between individual phase constituents of the aqueous phase involved in the forming reaction of this solid phase: Neither the solubility constant nor any of the involved Pitzer parameters must be changed individually. If the editor tries to do so, the system will prevent this and issue an appropriate note. If the set contains data from different editors, such a conflict must be clarified by personal communication between the editors.

Assignments can only be issued for an existing set. Each set is associated with a given interaction model.

Further it is possible to define relationships between sets. The relationship between two sets is either additive (AND) or exclusive (XOR). The former means, that two sets must always be used together while the latter means the contrary: either one or the other data set should be selected for the export into a parameter file.

10. Auditing

Many records in THEREDA are subject to a review process called “auditing”. The audit process is of paramount relevance for the quality assurance in THEREDA. It is formalized, controlled by internal functions directly incorporated into the DB-Control interface (see Section 11) and fully documented. Technically any data record with the “CalcMode” “entered” (see Section 6.3) is subject of an audit. The audit process is highly automated and based on both internal checks/calculations and an action scheme relying on “auditors”. The “auditor” checks the data and fills interactive forms. Depending on the “auditor’s” perception an appropriate status flag is assigned to the data record and thus marked for a data release or a revision. The flag may require further actions from the original data record editor, afterward initiating a new audit cycle.

11. DB-Control

Editors work with the database using a graphical user interface which can be accessed from the common project web site. This JAVA-based program is referred to as “DB-Control”.

Working with DB-Control actually spares the editor to deal with many separate tables, as these are reduced to lists or are combined to give useful views. Moreover, most input is checked by DB-Control immediately for formal correctness. Also the input sensibility is checked if possible.

Some more details can be found in the [supplements](#).

12. Export

The primary purpose of THEREDA is to provide users with ready-to-use parameter files. As THEREDA is a long-term project, and nobody can tell which geochemical codes will be used in the future, it was decided that the principal export format should be generic and not bound to any existing code at all. This generic output should encourage users to design own programs to import data from THEREDA. To ensure that the generic output really works it is further used as input to the creation of parameter files. Four codes for thermodynamic equilibrium calculations are supported at present: EQ3/6, PHREEQC, Geochemist’s Workbench[®], and ChemApp. All export files are provided with a MD5-checksum.

The generic output is written in JSON-format. JSON (JavaScript Object Notation) is a lightweight text-based open standard designed for human-readable data interchange. Although JSON is based on a subset of the JavaScript scripting language and is commonly used with that language, it is a language-independent data format. Code for parsing and generating JSON data is readily available for a large variety of scripting languages. The projects website <http://json.org> (accessed 2014-04-08) provides a comprehensive listing of existing and well-tested JSON libraries.

JSON is a generic plain text (ASCII) format with a low overhead. This combined with the above listed properties renders it a suitable tool to both provide long-term archiving of the database and to produce a well-structured yet human readable copy of the data-base (or parts of it). The latter is important to provide a generic, intermediate format for database content to be further processed, e.g., in various parsers/converters to produce input files specifically formatted for a variety of geochemical speciation codes.

The THEREDA database is designed as one main JSON object which is divided into four arrays (‘Elements’, ‘Phases’, ‘InteractionPhases’, and ‘Bibliography’). The order of these arrays is a THEREDA convention.

```

{                                     # main object opening
  'Elements':                          # opening array: "Elements"
  [                                     # content
    # ...                               # closing array
  ],
  'Phases':                             # opening array: "Phases"
  [                                     # content
    # ...                               # closing array
  ],
  'InteractionPhases'                 # opening array:
  [                                     # "InteractionPhases"
    # ...                               # content
  ],
  'Bibliography'                     # opening array:
  [                                     # "Bibliography"
    # ...                               # content
  ]
}                                     # closing array
}                                     # main object closing

```

More details about the JSON-structure used in THEREDA are given in the [supplements](#). A full documentation of the JSON structure applied to the THEREDA database is available for download on the THEREDA website. Thus external third-parties shall be enabled to write their own parsers/converters to address their specific needs, e.g. other in-house codes. In this spirit, any proposals and remarks, which help to improve our JSON export format for THEREDA, are most welcome.

13. Further system functionalities

There are some other functionalities which the editor hardly notice, and which do not deal directly with the actual database, but fulfil some important tasks nevertheless. They shall be summarized below.

The program "Table Log" (<http://pgfoundry.org/projects/table-log>) records changes within the database tables and thus provides transparency to all changes ever made. Additionally, it opens up the possibility to reset the THEREDA database on any previous state. Not only changes to records are saved, but also the insertion of new records as well as the deletion of records. Therefore it is possible to recover even deleted records.

For security reasons the whole databank including clone-copies for external access (released data), all scripts, auxiliary programs (export, logging) are mirrored daily on different servers.

14. Benchmarking with other database projects

In general, benchmarking with other databases is certainly an important issue. For example, another Pitzer database was developed within the frame of the Yucca Mountain Project (Mariner, 2004). However, we are convinced that "Benchmarking" is not meaningful as long as the values of entered or internally calculated thermodynamic data are compared. Rather, such a test should refer to the reproduction of published, well founded experimental data. Currently, efforts are undertaken in the THEREDA project, to establish such an evidence-based approach. Basically, it is our vision to provide CSV-formatted text files containing experimental data along with code-specific scripts, to enable users, both from the THEREDA project or other database projects, to validate their database with specific test cases. Failure to reproduce experimental results, primarily solubilities, will either indicate procedural

mistakes occurring during data capture and – management, or toward real gaps in the thermodynamic database.

There will be graphical representations for calculations with the THEREDA database only. However, it is hoped that such an instrument would encourage other partners to put their database to the test and initiate an informal exchange of knowledge for the mutual benefit of all.

15. Conclusions

THEREDA constitutes a web-based suite of codes designed to facilitate networked development and administration of a thermodynamic database. While the present focus lies on the development of a database optimized for solubility calculations in high-saline solutions, the same architecture could be used for the simultaneous development for more databases, e.g. for low ionic strength media. THEREDA sets priorities on the documentation of thermodynamic data with respect to their reliability and origin and strives to provide the user with ready-to-use parameter files. A generic data format is provided to promote wider usage of the data in THEREDA. This format could also be used as a common data format for the exchange between different database projects.

Acknowledgements

The development of THEREDA has been jointly funded by the German Federal Ministry of Education and Research (BMBF), the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMU), and the Federal Ministry for Economic Affairs and Energy (BMWi).

Gratitude is expressed towards the following colleagues who – together with the authors – contributed or are contributing to the development of THEREDA (in alphabetical order): Marcus Altmaier (KIT-INE), Christiane Bube (formerly KIT-INE), Xavier Gaona (KIT-INE), Sven Hagemann (GRS), Horst-Jürgen Herbert (GRS), Bernhard Kienzler (KIT-INE), Volker Neck † (formerly KIT-INE), Steffen Leske (formerly HZDR-IRE, still active), Anke Richter (HZDR-IRE), Tina Schrage (GRS), Dmitry Sukhanov (formerly TUBAF), Wolfgang Voigt (TUBAF), Stefan Wilhelm (AFC), Thomas Wilms (formerly GRS), Georgia Wollmann (formerly TUBAF).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.apgeochem.2014.12.016>.

References

- Altmaier, M. et al., 2014. "THEREDA: Thermodynamic Reference Database, Database for Radionuclides", Final Report BMWi Project THEREDA, INE Report 002/14.
- Bernarducci, E.E. et al., 1979. Partial molal heat capacity of aqueous ferrous chloride from measurements of integral heats of dilution. *J. Solution Chem.* 8 (10), 717–727. <http://dx.doi.org/10.1007/BF00648777>.
- Bethke, C.M., 2008. *Geochemical and Biogeochemical Reaction Modeling*. Cambridge University Press, Cambridge, ISBN-13: 978-0-521-87554-7.
- Bond, K.A. et al., 1997. "HATCHES: A referenced thermodynamic database for chemical equilibrium studies", Nirex Report NSS/R379.
- Bradbury, M.H., Baeyens, B., 1997. A mechanistic description of Ni and Zn sorption on Na-montmorillonite, Part II: Modelling. *J. Contam. Hydrol.* 27, 223–248. [http://dx.doi.org/10.1016/S0169-7722\(97\)00007-7](http://dx.doi.org/10.1016/S0169-7722(97)00007-7).
- Duro, L. et al., 2006. "Update of a thermodynamic database for radionuclides to assist solubility limits calculation for performance assessment" Swedish Nuclear Fuel and Waste Management Company, SKB Technical Report TR-06-17, ISSN 1404-0344.
- Duro, L. et al., 2012. ThermoChimie, the ANDRA thermodynamic database. In: *MRS Proceedings* 1475. doi: <http://dx.doi.org/10.1557/opl.2012.637>.
- Dzombak, D.A., Morel, F.M.M., 1990. *Surface Complexation Modeling. Hydrous Ferric Oxide*. Wiley, New York, ISBN-10: 0-471-63731-9.

- Eriksson, G. et al., 1995. A General Thermodynamic Software Interface. In: Proceedings 2nd Colloquium on Process Simulation, Helsinki University of Technology, 113 (Report TKK-V-B104).
- Eriksson, G. et al., 1997. ChemApp – a programmable thermodynamic calculation interface. In: Werkstoffwoche '96, Symposium 8 Simulation, Modellierung, Informationssysteme, 47, ISBN 3-88355-236-4.
- Frenkel, M. et al., 2006. XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture (ThermoML). *Pure Appl. Chem.* 78 (3), 541–612. <http://dx.doi.org/10.1351/pac200678030541>.
- Fuger, J. et al., 2008. *Chemical Thermodynamics, Chemical Thermodynamics of Thorium*, vol. 11. OECD Nuclear Energy Agency, Paris, ISBN-10: 926405667X, ISBN-13: 978-9264056671.
- Giffaut, E. et al., 2014. Andra thermodynamic data for performance assessment: *ThermoChimie. Appl. Geochem.* 49, 225–236. <http://dx.doi.org/10.1016/j.apgeochem.2014.05.007>.
- Grivé, M. et al., 2008. Assessment of the Concentration Limits for Radionuclides for Posiva. Posiva OY Working Report 2007-103.
- Hummel, W. et al., 2002. *Nagra/PSI Chemical Thermodynamic Data Base 01/01*. Universal Publishers, ISBN-13: 978-1581126204, ISBN-10: 1581126204.
- Hummel, W. et al., 2005. *Chemical Thermodynamics of Compounds and Complexes of U, Np, Pu, Am, Tc, Se, Ni and Zr with Selected Organic Ligands. Chemical Thermodynamics*, vol. 9. Elsevier, ISBN-13: 978-0-444-51402-8, ISBN-10: 0-444-51402-3.
- Jove-Colon, C. et al., 2007. Pitzer database development: Description of the Pitzer geochemical thermodynamic database data0.ypf.R2. Appendix I. In: In-Drift Precipitates/Salts Model, Report ANL-EBS-MD-000045 REV 03, DOC.20070306.0037 Las Vegas, Nevada: Sandia National Laboratories.
- Kitamura, A. et al., 2012. Update of JAEA-TDB: Additional Selection of Thermodynamic Data for Solid and Gaseous Phases on Nickel, Selenium, Zirconium, Technetium, Thorium, Uranium, Neptunium Plutonium and Americium, Update of Thermodynamic Data on Iodine, and Some Modifications. JAEA-Data/Code 2012-006.
- Kulik, D.A. et al., 2013. GEM-Selektor geochemical modeling package: revised algorithm and GEMS3K numerical kernel for coupled simulation codes. *Comput. Geosci.* 17 (1), 1–24. <http://dx.doi.org/10.1007/s10596-012-9310-6>.
- Mariner, P., 2004. In-Drift Precipitates/Salts Model. ANL-EBS-MD-000045 REV 02. November 2004. Prepared for: U.S. Department of Energy Office of Civilian Radioactive Waste Management, Office of Repository Development 1551, Hillshire Drive, Las Vegas, Nevada 89134-6321. Prepared by: Bechtel SAIC Company, LLC, 1180 Town Center Drive, Las Vegas, Nevada 89144. Contract Number DE-AC28-01RW12101.
- Mompeán, F.J., Wanner, H., 2003. The OECD nuclear energy agency thermochemical database project. *Radiochim. Acta* 91 (11), 617–622. <http://dx.doi.org/10.1524/ract.91.11.617.23468>.
- Parkhurst, D.L., Appelo, C.A.J., 2013. Description of input and examples for PHREEQC version 3—a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. In: U.S. Geological Survey Techniques and Methods, book 6, chap. A43, p. 497. <<http://pubs.usgs.gov/tm/06/a43/>>.
- Pitzer, K.S., 1991. Ion interaction approach: theory and data correlation. In: Pitzer, K.S. (Ed.), *Activity Coefficients in Electrolyte Solutions*, second ed. CRC Press, Boca Raton, Florida, pp. 75–153.
- Schindler, P.W., Gamsjäger, H., 1972. Acid-base reactions of the TiO₂ (Anatase) – water interface and the point of zero charge of TiO₂ suspensions. *Kolloid Z.* 250, 759–763. <http://dx.doi.org/10.1007/BF01498568>.
- Thoenen, T. et al., 2004. Development and application of the Nagra/PSI Chemical Thermodynamic Data Base 01/01. *Geol. Soc. Lond. Spec. Publ.* 236 (1), 561–577. <http://dx.doi.org/10.1144/GSL.SP.2004.236.01.31>.
- Wagner, T. et al., 2012. GEM-Selektor geochemical modeling package: TSoLMol library and data interface for multicomponent phase models. *Can. Mineral.* 50 (5), 701–723. <http://dx.doi.org/10.3749/canmin.50.5.1173>.
- Wang, L. et al., 2010. MOLDATA: A Thermochemical Data Base for phenomenological and safety assessment studies for disposal of radioactive waste in Belgium. SCK•CEN-ER-121, SCK•CEN, Mol, Belgium.
- Westall, J.C., Hohl, H., 1980. A comparison of electrostatic models for the oxide/solution interface. *Adv. Colloid Interface Sci.* 12 (4), 265–294. [http://dx.doi.org/10.1016/0001-8686\(80\)80012-1](http://dx.doi.org/10.1016/0001-8686(80)80012-1).
- Wolery, T., 1992. EQ3/6, A software package for the geochemical modelling of aqueous systems. UCRL MA 110662 Part I. Lawrence Livermore National Laboratory.
- Yoshida, Y., Shibata, M., 2004. Establishment of Data Base Files of Thermodynamic Data developed by OECD/NEA Part II – Thermodynamic data of Tc, U, Np, Pu and Am with auxiliary species, JNC Technical Report, JNC TN8400 2004-025.