A Crystal Structure Database for the Elements and Select Diatomic Compounds

Carlos Cruz
Aamer Haque
Sandy Landsberg
Thoraya Zedan

George Mason University
Institute for Computational Sciences and Informatics
School of Information Technology and Engineering

Overview

- Project Description
- Background and Technical Challenges
- Crystal Data
- Database Specifications
- User Model
- Web Interface
- Prototype Database with Demonstration
- Summary
- Future Work

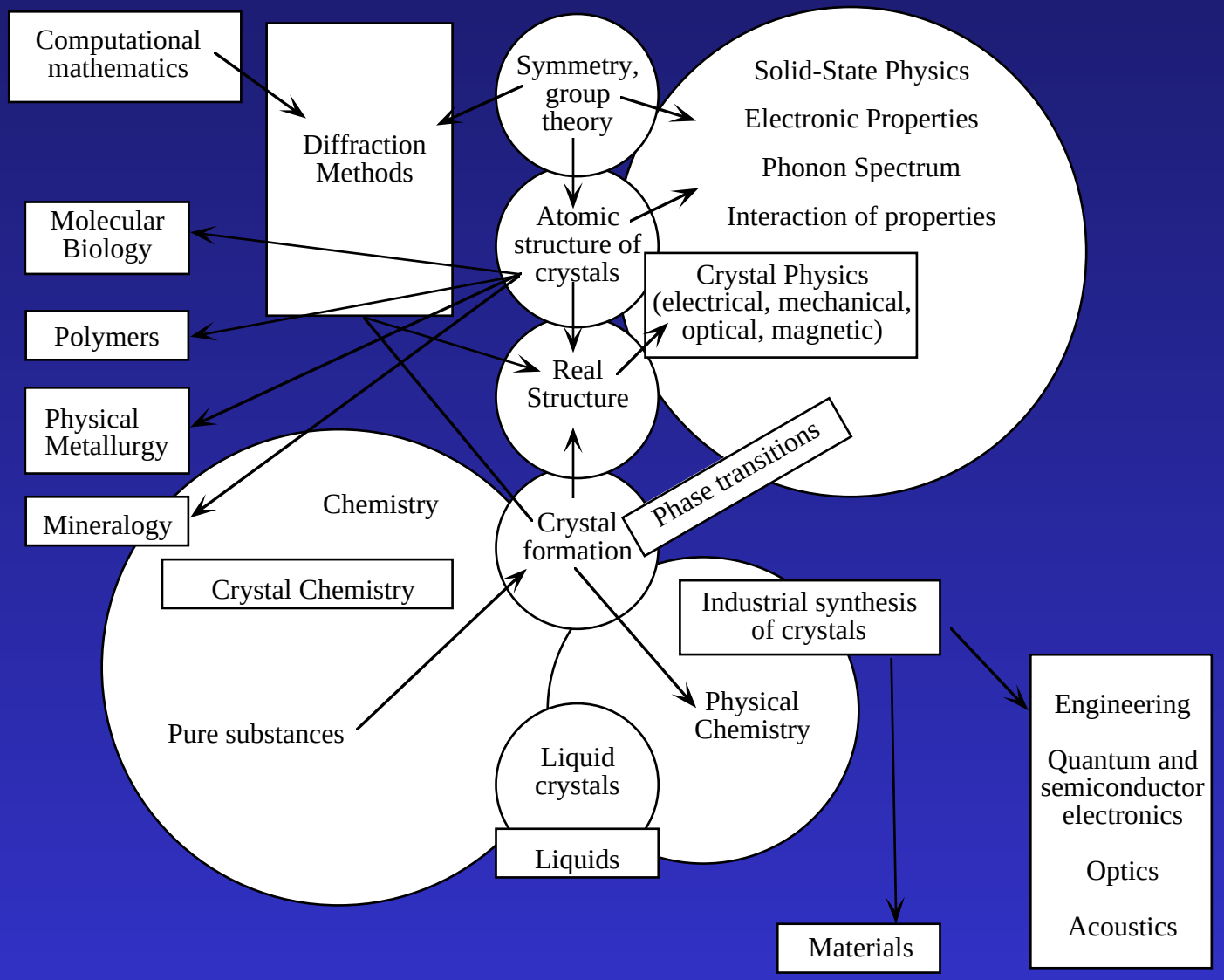
Project Description

- Elements and RX Diatomic Compounds
- Designed for material scientists and solid-state physicists
- Available Crystal Databases:
  - NIST Crystal Data
  - WebElements
- Crystal Structure Reference Texts:
  - Kittel
  - Wyckoff
  - Hahn
  - Donnay

Project Description

- Primary Objectives:
  - Provide a database management system for inorganic crystal structure data
  - Provide a database model extendable to more complex compounds
  - Design an easy-to-use graphical user interface for query generation and execution
  - Provide Internet access via the Web to an Oracle database
  - Provide textual and 2-D graphics output based on queries (via the Web)

Background

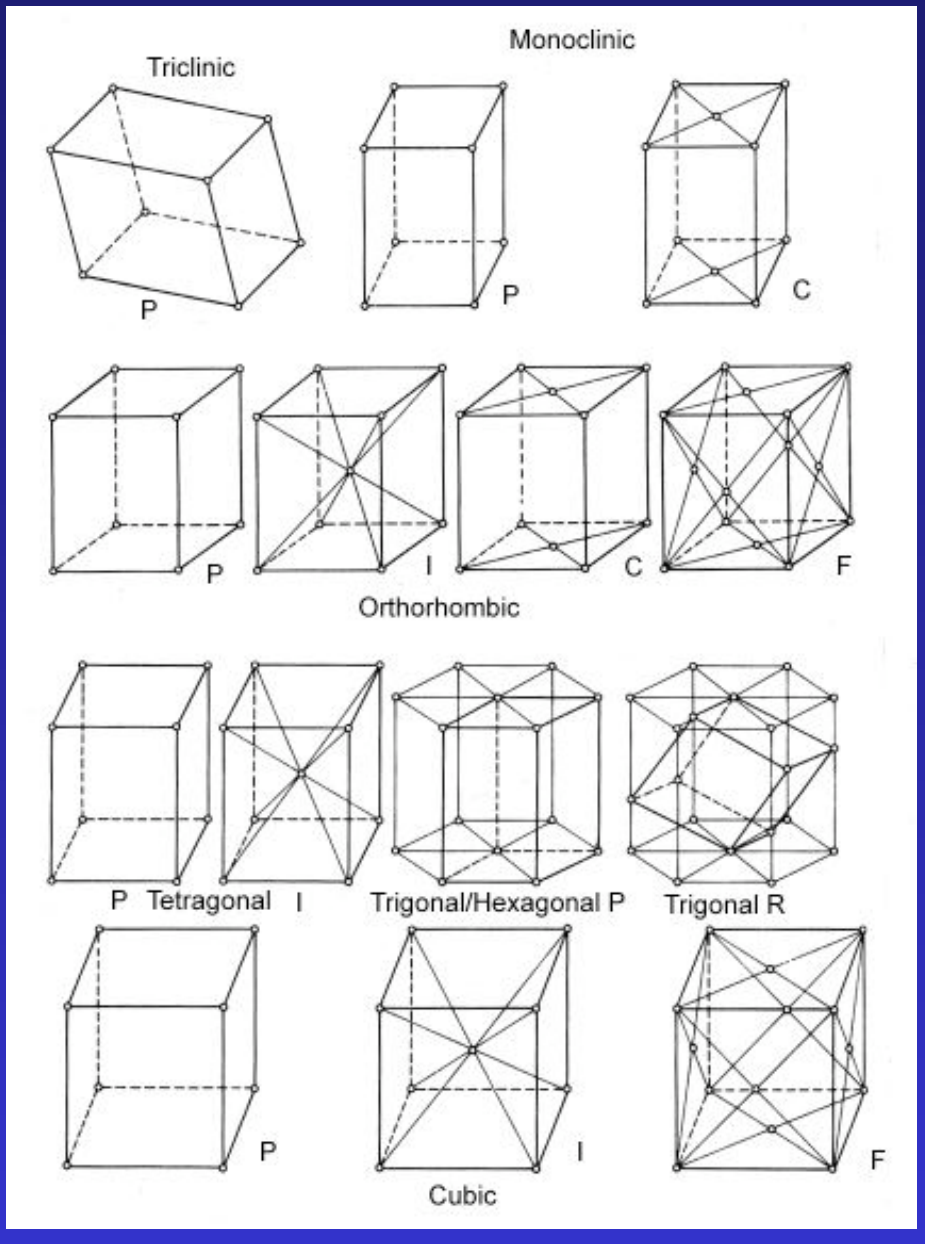


Branches of Crystallography and its Relation to Other Sciences

The Basics of Crystallography

7 Crystal Systems	32 Point Groups
14 Bravais Lattice Types	230 Space Groups

Crystal System	Point Group
Triclinic	1, -1
Monoclinic	2, m, 2/m
Orthorhombic	222, mm2, mmm
Tetragonal	4, -4, 4/m, 4222, 4mm, -42m, 4/mmm
Trigonal	3, -3, 32, 3m, -3m
Hexagonal	6, -6, 6/m, 622, -62m, 6/mmm
Cubic	23, m-3, 432 -43m, m3m



Technical Challenges

- Determination and accessibility of relevant data
  - Key data
  - Physical data
  - Crystallographic data
- Organization of data into an entity-relationship diagram
  - quick cross-correlations
  - extendable to more complex compounds
- Data Fusion
- NIST Data Decoding

Technical Challenges

- Database Implementation
- Graphical User Interface Design and Implementatoin
- Dynamic Plotting
- Oracle and the Web
- Extracting New Science

Crystal Data

- Data Description
- Quality of Data
- NIST Data Decoding
- Representative Queries
- Representative Plotting

Data Description

- Key Data (E/D)
  - Symbol (E/D)
  - Name (E/D)
  - Element Group (E)
  - Element Period (E)

Data Description

- Physics Data
  - Atomic Number (E)
  - Mass Properties (E/D)
    - Molecular Weight (E/D)
    - Density (E/D)
  - Temperature (E/D)
    - Boiling Pt
    - Melting Pt
  - Radii (E)
    - Standard radii for ions in inert gas configuration
    - Radii of atoms when in tetrahedral covalent bonds
    - Radii of ions in 12-coordinated metals
  - Ionization Energies (E)
    - One Electron
    - Two Electron

Data Description

- Crystallographic Data
  - Unit cell data (E/D)
    - Lattice parameters (a, b, c, α,β, γ, volume)
    - System Structure (cubic, hexagonal, triclinic, . . .)
      - Structure type (e.g. cubic types: sc, fcc, bcc)
      - Structure info
    - Point group
    - Space group
  - Diatomic Structure type (e.g. NaCl type, CsCl type, . . .) (D)
  - Crystal Properties (E/D)
    - Bulk Modulus
    - Nearest Neighbor Distance
    - Cohesive Energy
  - Vibrational lattice parameters (E/D)

Quality of Data

- NIST Crystal Data
  - Chemical, physical, and crystallographic information
  - 210,403 inorganic and organic crystalline materials
  - Standard cell parameters, cell volume, space group number and symbol, density, and classification by chemical type, chemical formula, and chemical name
  - Meta-data: associated literature reference
  - Each element and/or compound can have several entries/records corresponding to different polymorphs or structures under special thermodynamic conditions
- Cross references to other sources, e.g. WebElements, Kittel, etc.

This will ensure consistency of the data with other sources.

NIST Data Decoding/Data Fusion

****

****

- Elements - Criterion for Consistency
  - Space group from NIST Crystal Data same as WebElements
  - Inconsistencies cross-referenced with the crystal structures given by Wyckoff
- Diatomic Compounds- Criterion for Consistency
  - all the possible characterizations for a compound were extracted
- NIST Data Decoding
  - format given in the NIST Crystal Data Database Specifications Manual
  - the crystal data cell information extracted from record type E
  - obeys the rules given by Donnay

Representative Queries

1. List the elements that have a melting point temperature greater than 1000K and less than 2000K.
2. List the elements that have a space group of P63/mmc with a hexagonal structure.
3. List the diatomic compounds with an NaCl structure and the lattice parameter, a.
4. List all the available information about the element Si.
5. List all the elements in Group 1.
6. List all the tetragonal structures with one-electron ionization energy greater than 5eV.
7. List the name, symbol and space group for all diatomic compounds with contain Ge.
8. List the standard atomic radii of the elements with a bcc structure.
9. List the possible combinations of RX compounds between elements in Group 1 and Group 17.
10. List all the possible RX compounds containing Si.

Representative Plotting

1. Plot melting point vs. atomic number for Period 4 elements.
2. Plot the lattice parameter, a, vs. melting temperature for elements with a cubic structure.
3. Plot the lattice parameter, a, vs. molecular weight for the CsCl diatomic compounds.
4. Plot the melting point temperature vs. atomic number for Group 1 and Group 2 elements.
5. Plot two-electron ionization energy vs. atomic number for the elements.
6. Plot density vs atomic number of Group 2 elements.
7. Plot cohesive energy vs. atomic number of all elements with cubic structure.
8. Plot bulk modulus vs. atomic number of all elements with a cubic structure.
9. Plot the cohesive energy vs. atomic number for Period 4 elements.
10. Plot lattice parameter, a, vs. molecular weight of all diatomic compounds that contain Aluminum and have lattice parameter, a > 3A.

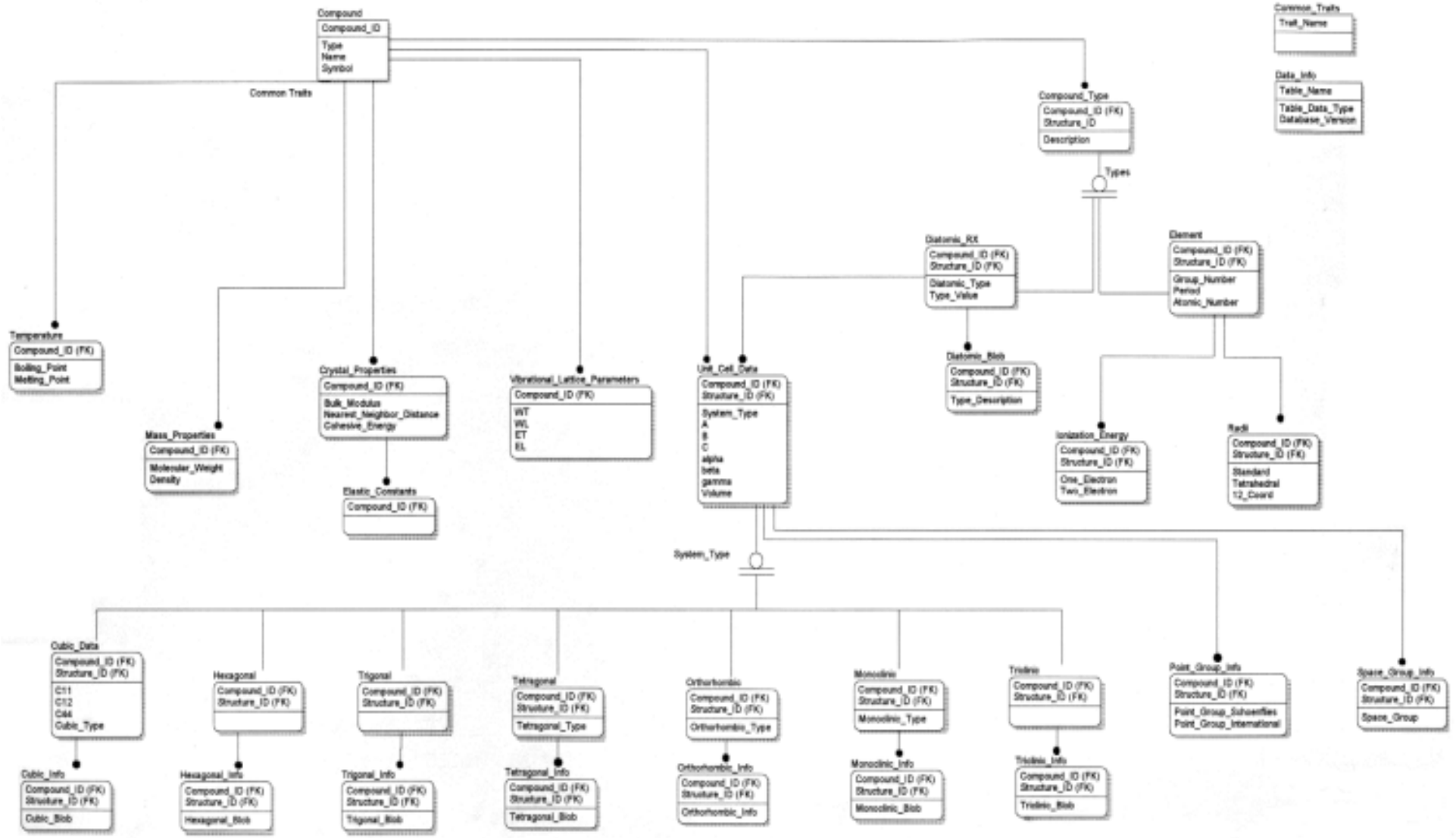
Database Specifications

- Database Rationale
- Conceptual Model
- Data Access Methods
- Database Software

Database Rationale

- Primary Key : NIST Compound Identifiers
- Compound Table Includes:
  - Elements
  - Diatomic Compounds
- Attributes Common to Both Branches of the Compound Table
- Attributes Specific to the Element and Diatomic RX Tables

Conceptual Model



Data Access Methods

- Boolean operations
  - distinct search parameters are selected
  - for each parameter, the database is searched and the subset of data consisting of the “hits” are saved
  - to answer the question, the subsets of data are intersected using Boolean “AND”, “OR” and “NOT” operations

Database Software

- ERWin/ERX 2.6.2.
- Oracle 7 Server Release 7.3.3
- SQL*Plus: Release 3.3.3.0.0
- Microsoft Visual Source Safe Version 5.0
- Designer/2000 and Developer/2000 1.3.2
- Oracle Webserver 2.1

User Model

- Interaction Requirements
- Allowable Usage
- Processing Steps
- Archival Strategy
- Curation Strategy
- Migration Strategy

Interaction Requirements

- Data classes:
  - Key data (text)
  - Physics data (numerical)
  - Crystal data-1 (numerical)
  - Crystal data-2 (text)
- Data Sources:
  - NIST Crystal Data (© 1997)
  - WebElements
  - Reference texts (Kittel, Wyckoff, etc.)
- Graphical User Interface: Web-based
- Network Access: Internet access

Allowable Usage

- Users: Novices, Expert, System Admin.
- Novices:
  - Students, All others
  - Web-based GUI provides easy data querying and plotting
- Expert Users:
  - Crystallographers, Chemists, Physicists, etc.
  - Web-based GUI
- System Admin.: Controls database tables
- No restrictions on audience size

Processing Steps

- Data Input:
  - electronic file transfer
  - manual data entry
- Data Fusion:
  - complex process
  - cross-check data against other available sources
- Production Generation:
  - ERWin used to forward engineer database
- Data Output:
  - text
  - 2-D Graphics
- Processing Requirements:
  - Workstation
- Storage Requirements:
  - currently 10 MB disk space (and Oracle User space)

Archival Strategy

- Data:
  - Tape backups
  - Data not dynamic
- Web Interface/other files:
  - Tape backups
- Oracle database/tables:
  - Oracle Data Manager

Curation Strategy

- Individual Domain-Based:
  - Current strategy
  - Two project members: Carlos Cruz and Sandy Landsberg
- Domain-Based Committee:
  - Used by Cambridge Crystallographic Data Centre for Cambridge Structural Database
  - Over 175,000 organic compounds
  - Standard form to submit new crystal data, Crystallographic Information File (CIF)

Migration Strategy

- Software:
  - Use established software (HTML, Oracle, Microsoft Access)
- Object-Oriented Model:
  - Current relational database design can transition to object-oriented model
  - Oracle 7.3.3 should work under Oracle 8
- Parallel-Processing:
  - Not necessary for current database
  - Possibly useful for querying over parameter space

Web Interface

- Web Security
- Unix-Based Dynamic Plotting
- Oracle Database Web Interface
- Oracle Dynamic Plotting

Web Security

- Common Gateway Interface (CGI)
- Processing Steps:
  - The CGI processes-server receives a request from a client.
  - A CGI program is executed.
  - The CGI program reads in the data, processes the data, and sends back output.
  - The CGI processes-server sends the output to the client.

Web Security

- User Authentication Process:
  - Get the entered user name and password from the input data.
  - Verify that the entered user name is the name of a valid user.
  - **If** the user name is valid,

    then salt the entered password using the stored random number and encrypt this password,

    else print out an error message.
  - Compare this password with the correct password.
  - **If** the passwords match,

    then print out the embedded HTML file,

    else print out an error message.

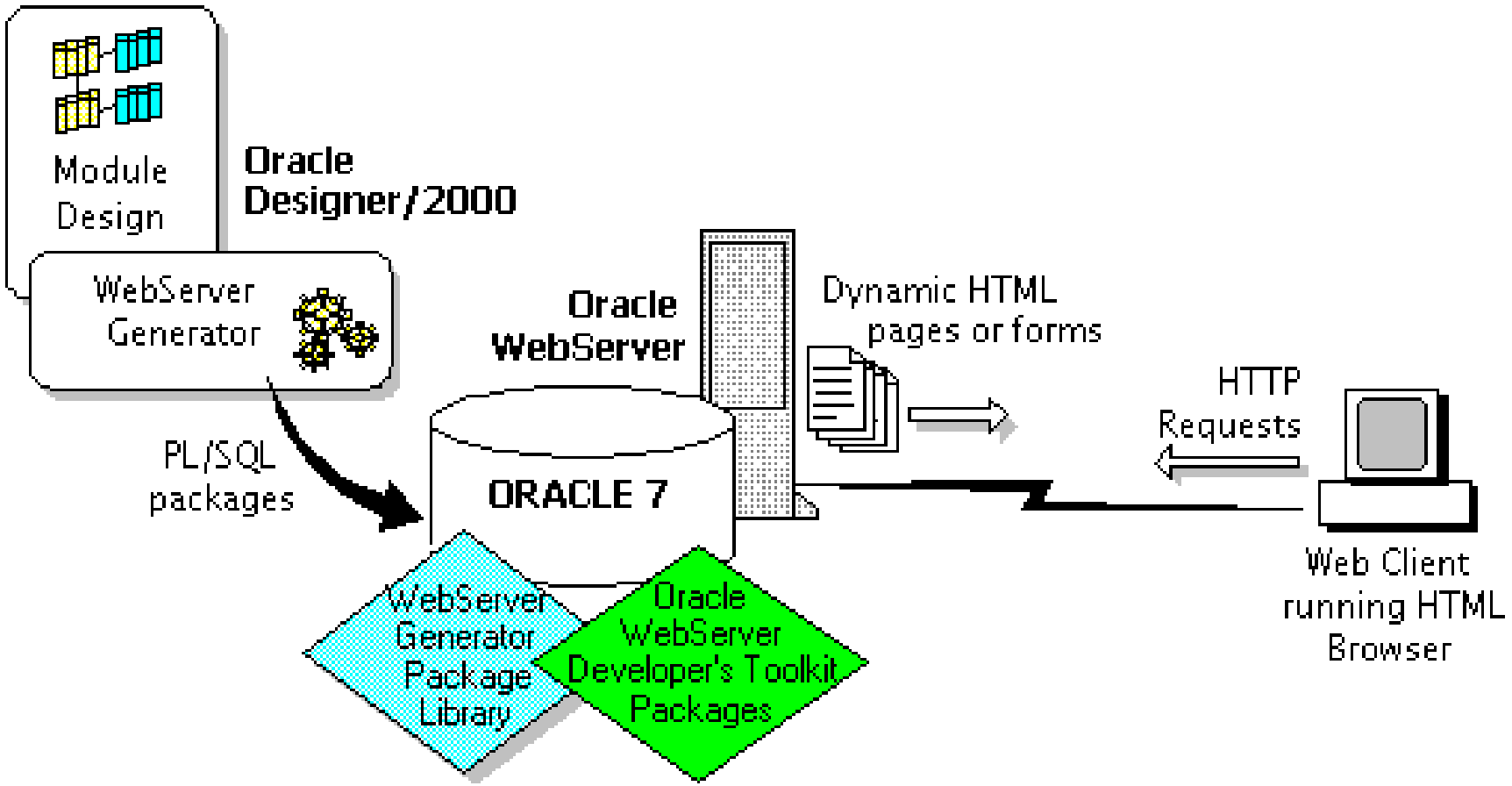
Unix-based Dynamic Plotting

- Dynamic Plotting using CGI and ‘gnuplot’:
- Processing steps:
  - The user submits the query on an HTML form.
  - The form submits the user input to a CGI program.
  - The CGI program gets the data for the plot from the database.
  - The CGI program writes out the data in X-Y table format to a file.
  - The CGI program executes the gnuplot program. The gnuplot program reads in the X-Y data from the file and generates a plot in postscript format.
  - The CGI program returns the postscript plot to the user.
- X-Y table format simply means that every line of the data file are as follows:
  - [X-coordinate][space][Y-coordinate][CR/LF]

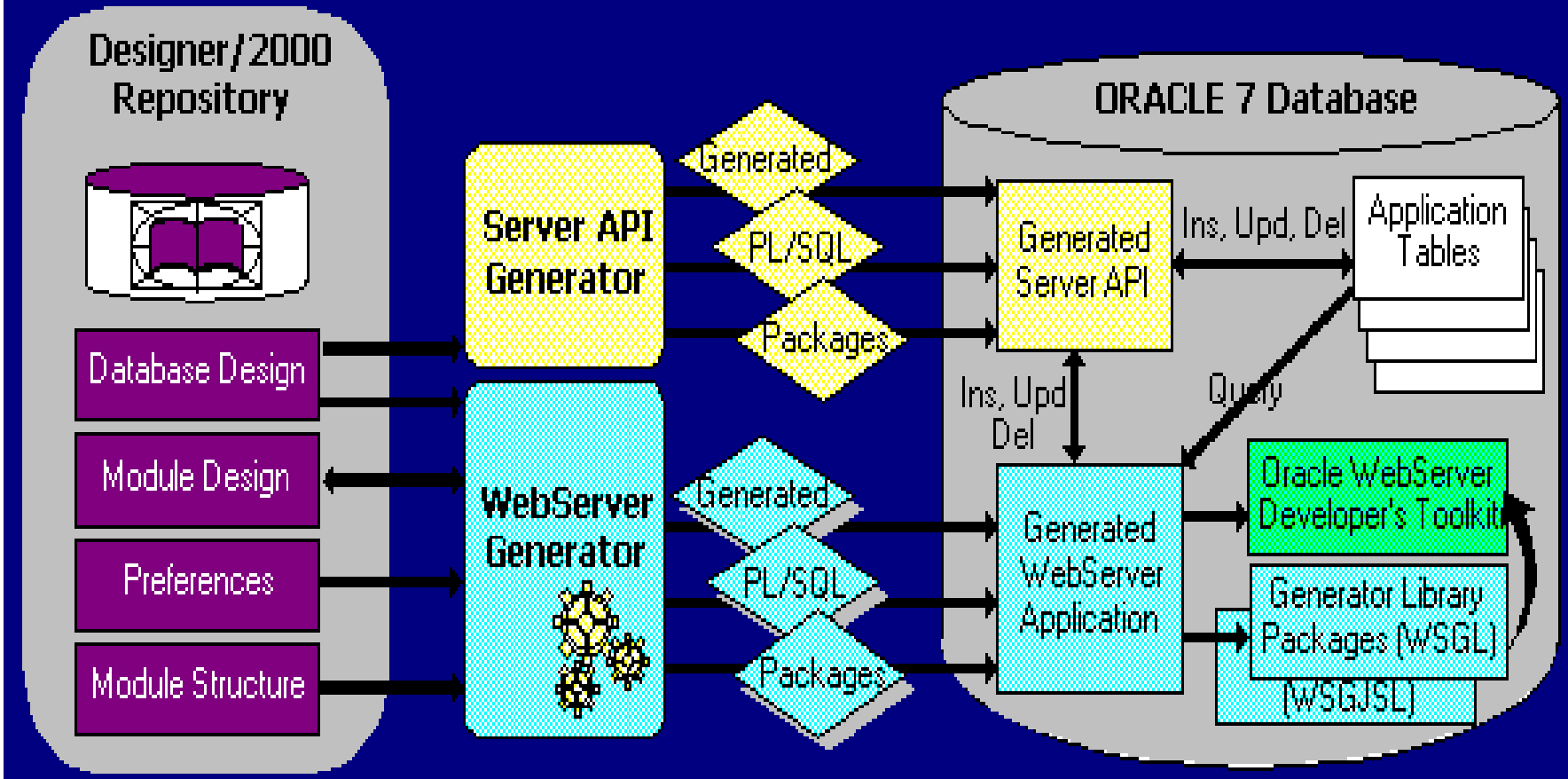
Oracle and The Web

- Connecting it all together:
  - Oracle Database
  - Oracle WebServer
  - Oracle Designer 2000

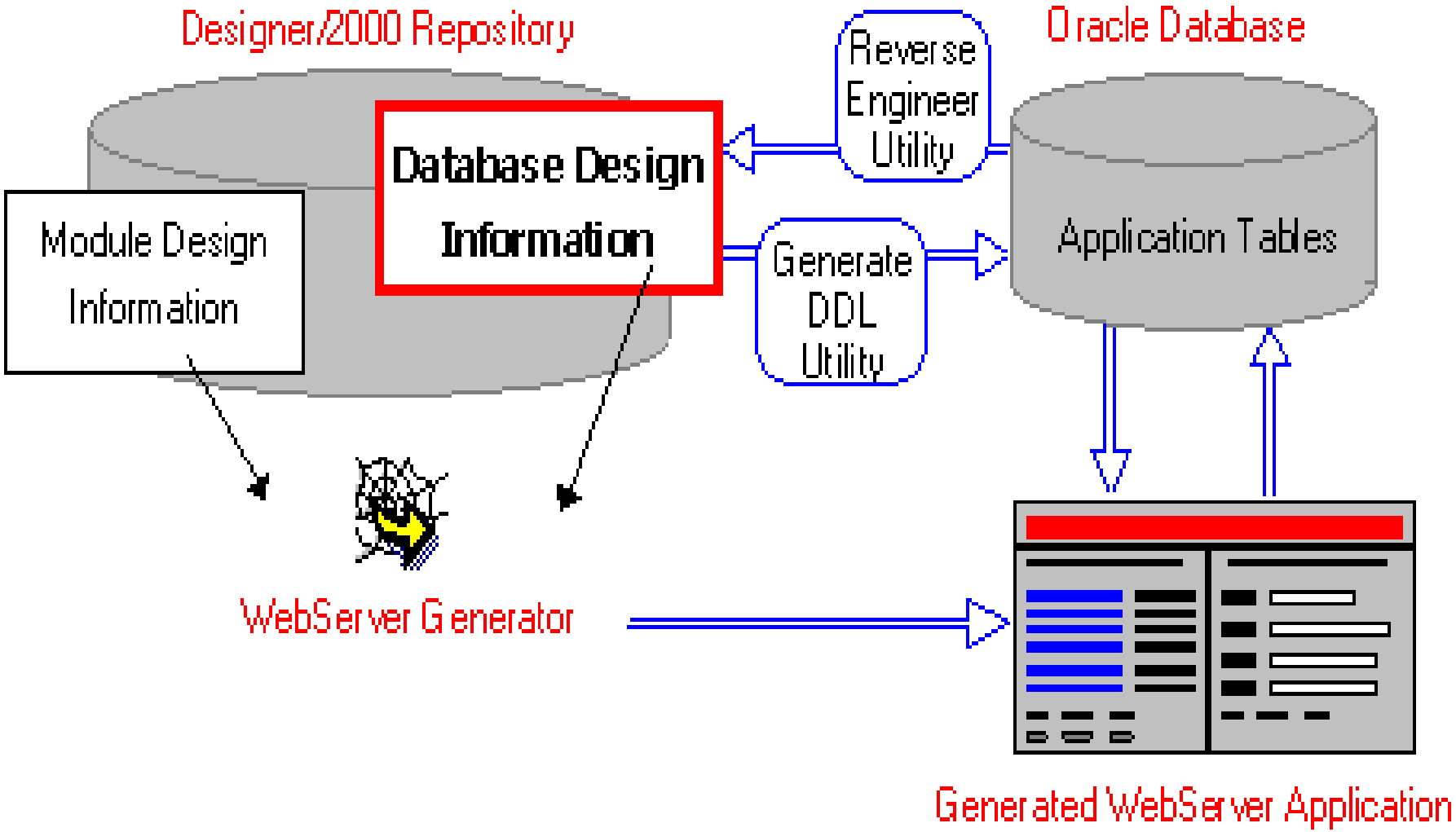
Introducing Designer/2000 WebServer Generator



How the WebServer Generator Works



Recording the Database Design



Setting up the WebServer User

- Set up: WebServer User
- Install: WebServer Developers Toolkit
- Install: WebServer Generator Library Packages
- Create: WebServer Database Connection Descriptors (DCD)
- Update: WebServer Web Request Broker.

Prototype Database

- Current Status
- Graphical User Interface
- Demonstration of Queries
  - Text Output
  - Plotting

Current Status - Element Data

- Type
- Name
- Symbol
- Group
- Period
- Atomic Number
- Boiling Point
- Melting Point
- Molecular Weight
- Ionization Energy
- Density
- Cell Volume
- Radii
- Unit Cell Data
- Space Group
- Bulk Modulus
- Cohesive Energy
- Crystal System

Current Status - Diatomic Compounds Data

- Type
- Name
- Symbol
- Molecular Weight
- Density
- Cell Volume
- Unit Cell Data
- Space Group
- Crystal System

Graphical User Interface

This applies to all demos shown:

****

COPYRIGHT NOTICE: Data from NIST Crystal Data (© 1997) are used with the permission of the National Institute of Standards and Technology.

****

Demonstration of Queries

Query 1. List the element that have a melting point temperature greater than 1000K and less than 2000K.

```
SELECT C.SYMBOL,
      T.MELTING_POINT
 FROM NIST_COMPOUND C, NIST_TEMPERATURE T
WHERE C.COMPOUND_ID = T.COMPOUND_ID
  AND T.MELTING_POINT >= 1000
  AND T.MELTING_POINT <= 3000 ;
```

SYMBOL	MELTING_POINT	SYMBOL	MELTING_POINT	SYMBOL	MELTING_POINT
Ru	2607	Am	1449	Ir	2739
Sc	1814	As	1090	Fe	1811
Si	1687	Ba	1000	La	1193
Ag	1235	B	2349	Mn	1519
Sr	1050	Ca	1115	Mo	2896
Tc	2430	Ce	1068	Nd	1297
Th	2115	Cr	2180	Ni	1728
Ti	1941	Co	1768	Ac	1323
U	1405	Cu	1358	Nb	2750
V	2183	Dy	1680	Pd	1828
Yb	1097	Er	1770	Pt	2041
Y	1799	Eu	1099	Pr	1208
Zr	2128	Gd	1585	Pa	1841
Sm	1345	Ge	1211	Rh	2237
Lu	1925	Au	1337		
Tb	1629	Hf	2506	48 rows selected.	
Tm	1818	Ho	1734	SQL>	

Demonstration of Queries

Query 2. List the elements that have a space group of P63/mmc with a hexagonal structure.

```
SELECT C.SYMBOL,
       E.ATOMIC_NUMBER,
       S.SPACE_GROUP,
       U.SYSTEM_TYPE
  FROM NIST_COMPOUND C,
       NIST_ELEMENT E,
       NIST_SPACE_GROUP_INFO S,
       NIST_UNIT_CELL_DATA U
 WHERE C.COMPOUND_ID = E.COMPOUND_ID
   AND E.COMPOUND_ID = S.COMPOUND_ID
   AND S.COMPOUND_ID = U.COMPOUND_ID
   AND U.SYSTEM_TYPE LIKE 'H%'
   AND S.SPACE_GROUP LIKE 'P63/mmc%'
 ORDER BY E.ATOMIC_NUMBER;
```

SYMBOL	ATOMIC_#	SPACE_GROUP	SYSTEM
He	2	P63/mmc	H
Be	4	P63/mmc	H
C	6	P63/mmc	H
N	7	P63/mmc	H
Mg	12	P63/mmc	H
Sc	21	P63/mmc	H
Ti	22	P63/mmc	H
Co	27	P63/mmc	H
Zn	30	P63/mmc	H
Y	39	P63/mmc	H
Zr	40	P63/mmc	H
Tc	43	P63/mmc	H
Ru	44	P63/mmc	H
Cd	48	P63/mmc	H
La	57	P63/mmc	H
Ce	58	P63/mmc	H
Pr	59	P63/mmc	H
Nd	60	P63/mmc	H
Gd	64	P63/mmc	H
Tb	65	P63/mmc	H
Dy	66	P63/mmc	H
Ho	67	P63/mmc	H
Er	68	P63/mmc	H
Tm	69	P63/mmc	H
Lu	71	P63/mmc	H
Hf	72	P63/mmc	H
Re	75	P63/mmc	H
Os	76	P63/mmc	H
Tl	81	P63/mmc	H
Am	95	P63/mmc	H

31 rows selected.
SQL>

Demonstration of Queries

Query 7. List the name, symbol and space group for all diatomic compounds with contain Ge.

```
SELECT SUBSTR(C.NAME, 1, 25) NAME, C.SYMBOL, S.SPACE_GROUP
  FROM NIST_COMPOUND C, NIST_SPACE_GROUP_INFO S
 WHERE C.COMPOUND_ID = S.COMPOUND_ID
   AND S.SPACE_GROUP != 'blank'
   AND C.SYMBOL LIKE 'Ge%'
   AND C.TYPE = 'DIATOMIC';
```

NAME	SYMBOL	SPACE_GROUP	NAME	SYMBOL	SPACE_GROUP
Germanium titanium	GeTi	P2mm	Germanium sulfide	GeS	Pbnm
Germanium ytterbium	GeYb	I4/***	Germanium sulfide	GeS	P6322
Germanium yttrium	GeY	Cmcm	Germanium sulfide	GeS	Pnam
Germanium yttrium	GeY	I4/***	Germanium sulfide	GeS	Pbnm
Germanium telluride	GeTe	Pnma	Germanium telluride	GeTe	R**
Germanium telluride	GeTe	R	Germanium telluride	GeTe	Fm3m
Germanium telluride	GeTe	R	Germanium telluride	GeTe	Fm3m
Germanium telluride	GeTe	R	Germanium telluride	GeTe	R3m
Germanium terbium	GeTb	Cmcm	Germanium telluride	GeTe	R
Germanium terbium	GeTb	I4/***	Germanium selenide	GeSe	Pbnm
Germanium thorium	GeTh	Fm3m	Germanium selenide	GeSe	P6322
Germanium thulium	GeTm	I4/***	Germanium selenide	GeSe	Pbnm
Germanium strontium	GeSr	Cmcm	Germanium sodium	GeNa	P21/c
Germanium strontium	GeSr	Cmcm	Germanium palladium	GePd	Pnam
Germanium strontium	GeSr	Cmcm	Germanium phosphide	GeP	C2/m

Demonstration of Queries

Query 7. List the name, symbol and space group for all diatomic compounds with contain Ge.

```
SELECT SUBSTR(C.NAME, 1, 25) NAME, C.SYMBOL, S.SPACE_GROUP
  FROM NIST_COMPOUND C, NIST_SPACE_GROUP_INFO S
 WHERE C.COMPOUND_ID = S.COMPOUND_ID
   AND S.SPACE_GROUP != 'blank'
   AND C.SYMBOL LIKE 'Ge%'
   AND C.TYPE = 'DIATOMIC';
```

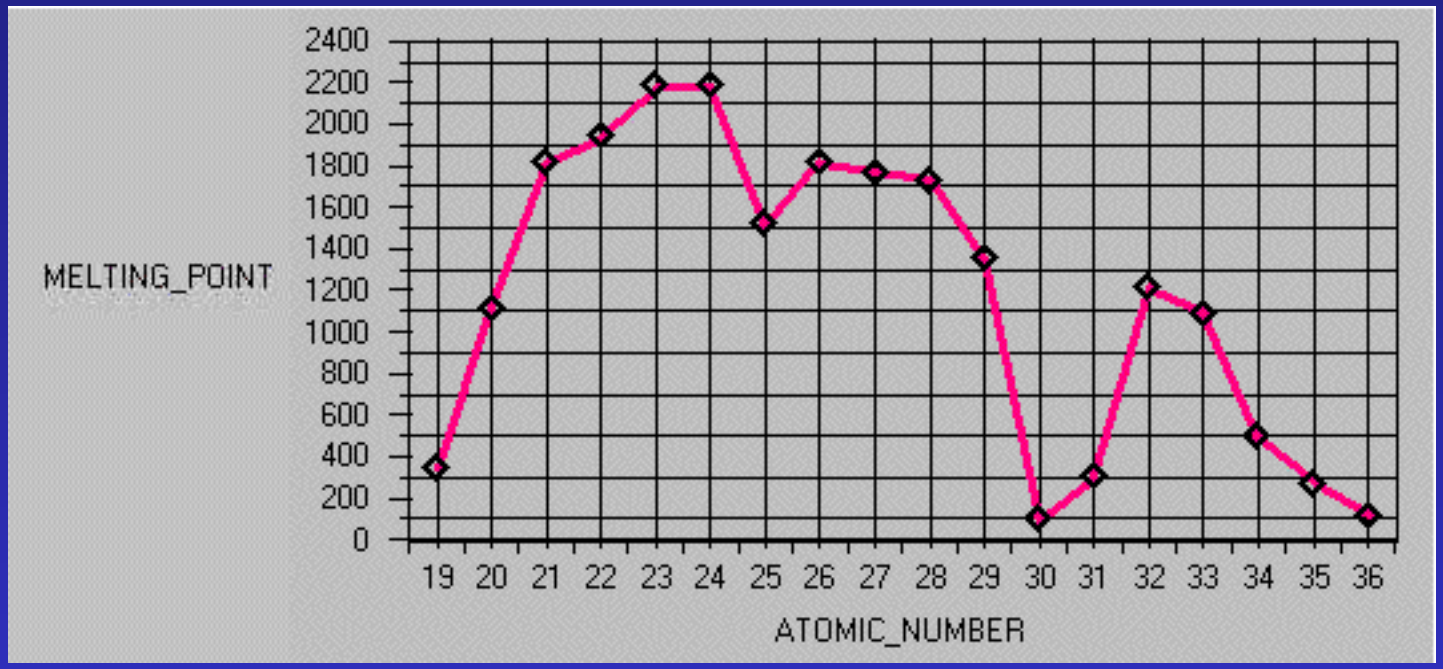
NAME	SYMBOL	SPACE_GROUP
Germanium platinum	GePt	Pnam
Germanium platinum	GePt	Pnam
Germanium praseodymium	GePr	Cmcm
Germanium praseodymium	GePr	Pbnm
Germanium praseodymium	GePr	Cmcm
Germanium rhodium	GeRh	Pnam
Germanium samarium	GeSm	Cmcm
Germanium samarium	GeSm	Cmcm
Germanium scandium	GeSc	Cmcm
Germanium selenide	GeSe	Pbnm
Germanium holmium	GeHo	Cmcm
Germanium holmium	GeHo	Cmcm
Germanium iridium	GeIr	Pnam
Germanium lanthanum	GeLa	Pbnm
Germanium lithium	GeLi	I41/a
Germanium lithium	GeLi	I41md
Germanium lutetium	GeLu	I4/***
Germanium lutetium	GeLu	R*
Germanium neodymium	GeNd	Cmcm
Germanium neodymium	GeNd	Cmcm
Germanium nickel	GeNi	Pnam

51 rows selected.

SQL>

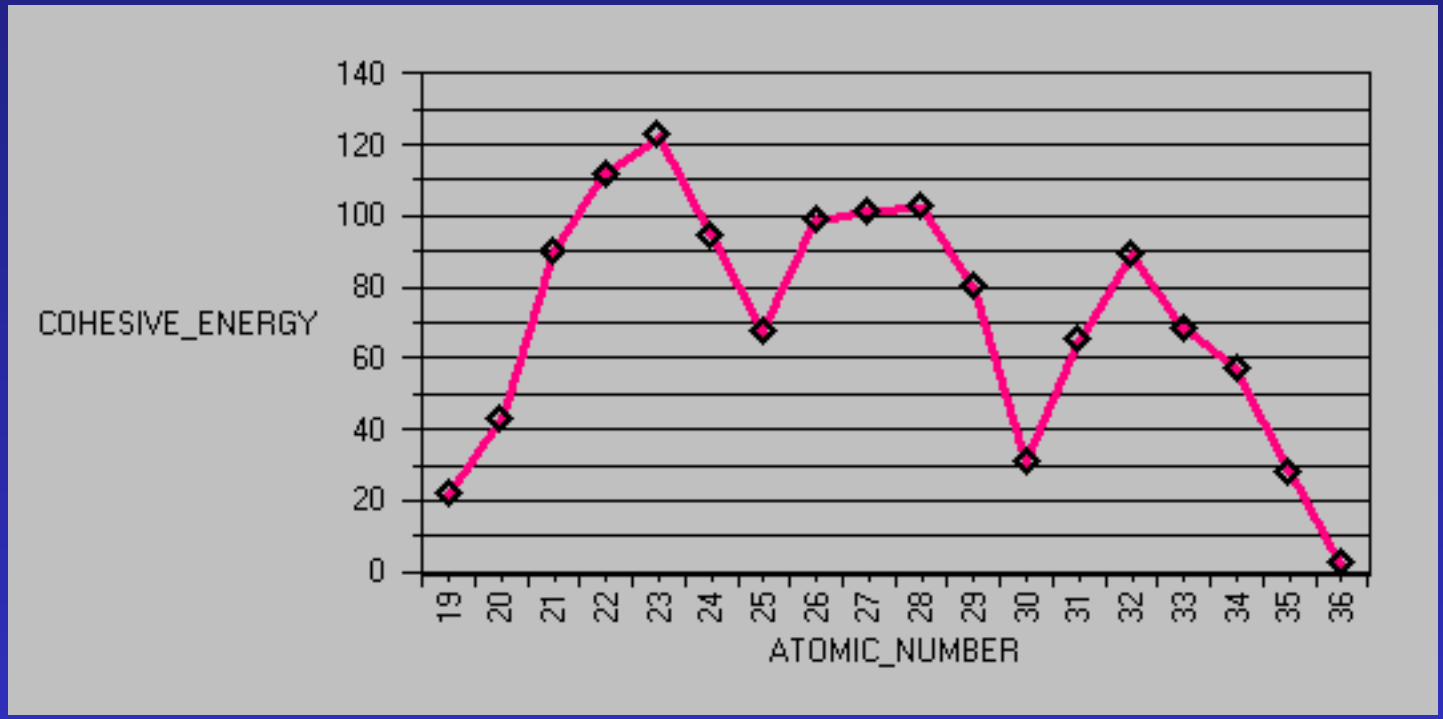
Demonstration of Plotting

Plot 1. Plot melting point versus atomic number for Period 4 elements.



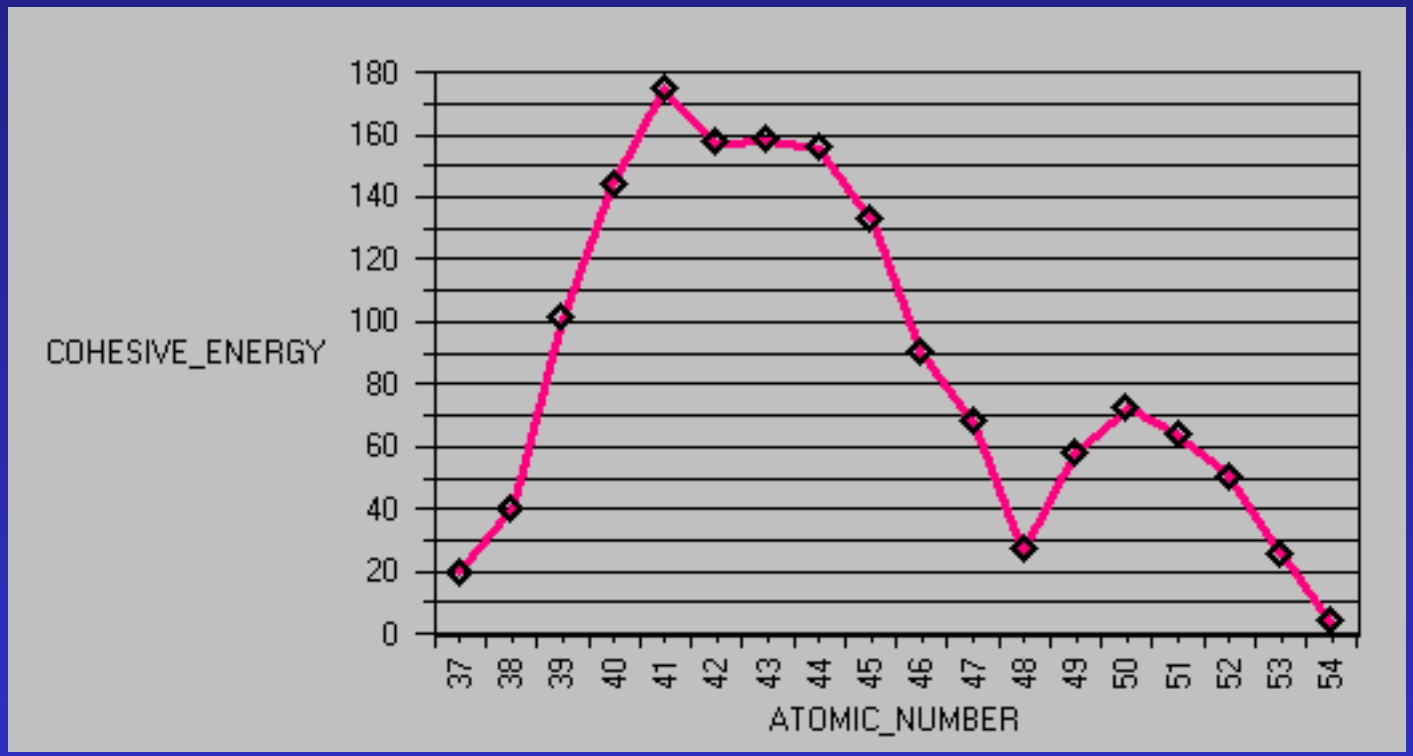
Demonstration of Plotting

New Plot. Plot cohesive energy versus atomic number for Period 4 elements.



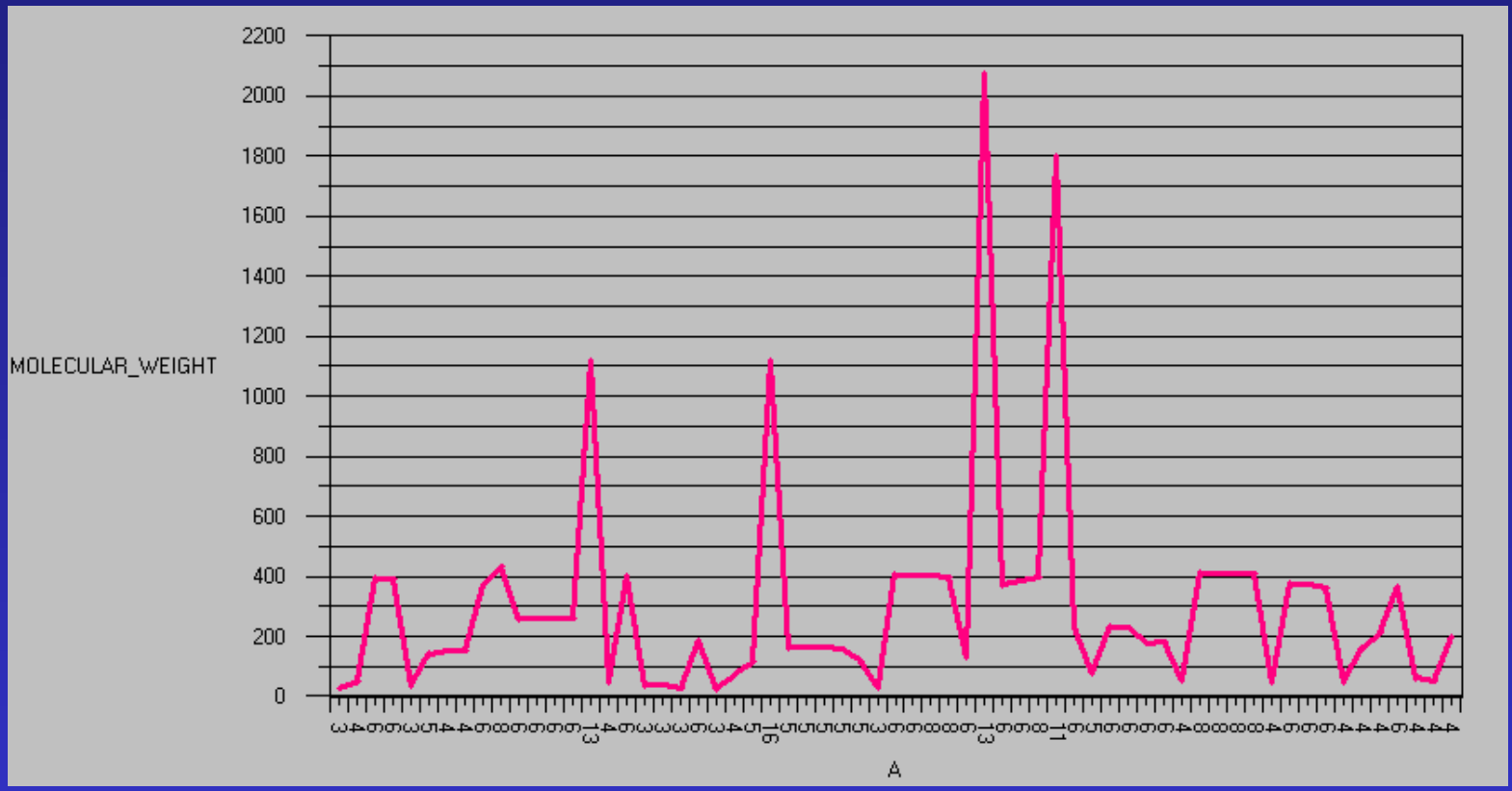
Demonstration of Plotting

New Plot. Plot cohesive energy versus atomic number for Period 5 elements.



Demonstration of Plotting

New Plot. Plot the lattice parameter, a, vs molecular weight of all diatomic compounds where a > 3A.



Summary

- Crystal data collection and fusion from multiple data sources
- Designed an entity-relationship schema for rapid data access, extendable to more complex compounds
- Implemented and demonstrated a working Oracle crystal structure database
- Provided a database management system for inorganic crystal data

Summary

- Designed and implemented a Web-based graphical user interface
- Demonstrated dynamic plotting using two methods
- Extensive work on connecting an Oracle database to the Web using Oracle Web Server and Designer/2000
- Generated queries and plots showing functionality of database as well as being scientifically relevant

Future Work

- Near-Term Goals:
  - Finish data collection and table population
  - Add associated meta-data
  - Complete and enhance graphical user interface
  - Fully-functioning Web interface to Oracle database connection
  - Fully-functioning dynamic plot generation

Future Work

- Long-Term Goals:
  - Extend the database to include other available complex inorganic compounds
  - Developing a Microsoft Access 97 interface and database
- New Crystallographic Studies:
  - Extract and add point group information, structure information
  - Query database based on point group, structure informatoin and crystal systems
  - Investigate physical and structural properties of rare-earth compounds

Acknowledgments

- NIST
  - Dr. Vicky Karen
  - NIST Standard Reference Data Program
- GMU
  - Prof. Estela Blaisten-Barojas
  - Profs. George Michaels and Larry Kerschberg