



COMPUTATIONAL **A**NALYSIS OF **S**Ocial AND **O**RGANIZATIONAL **S**YSTEMS

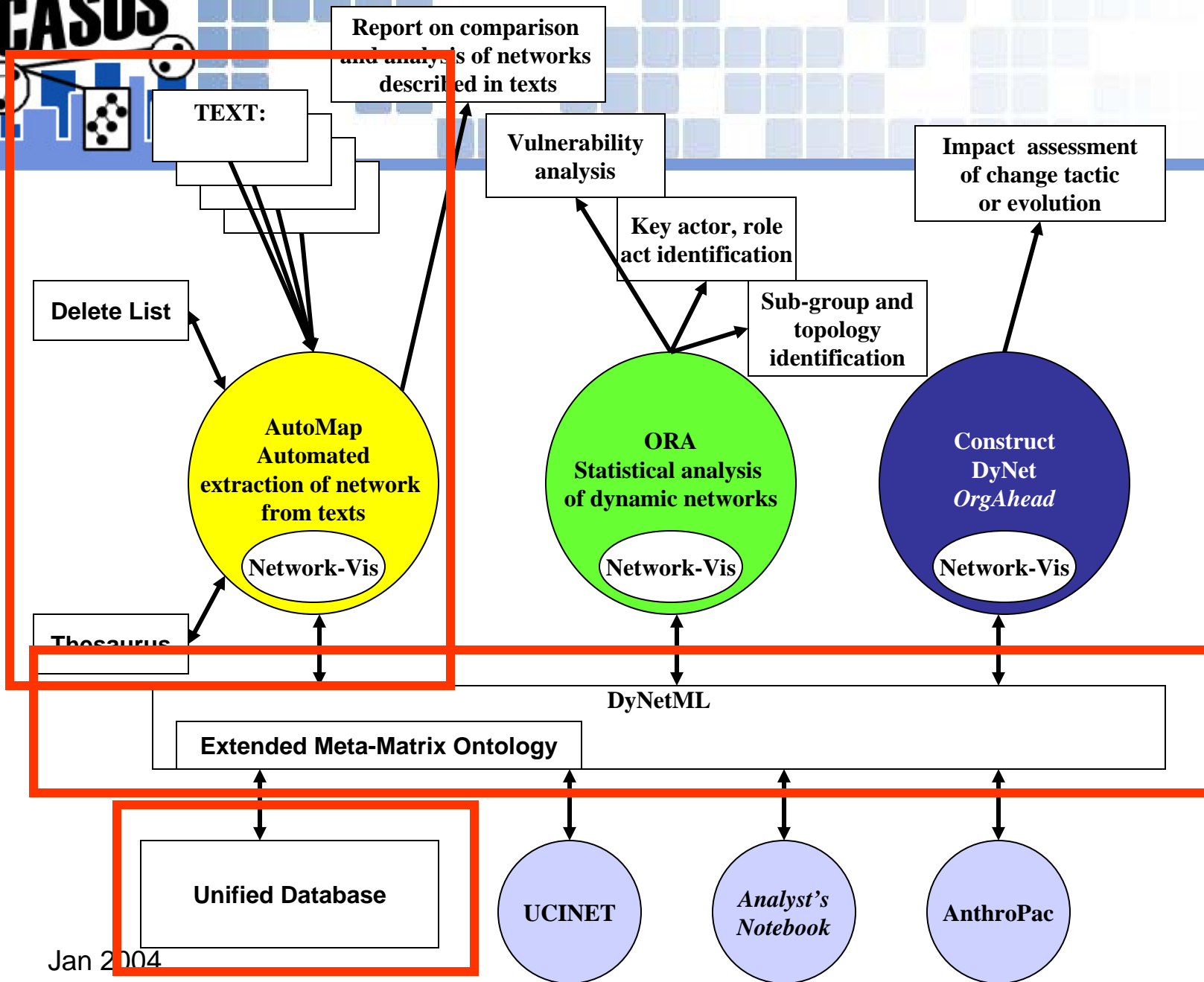


Data Integration and Manipulation

Social Network Databases

DyNetML

Analysis Tool Integration



CASOS



Social Network Data and Databases



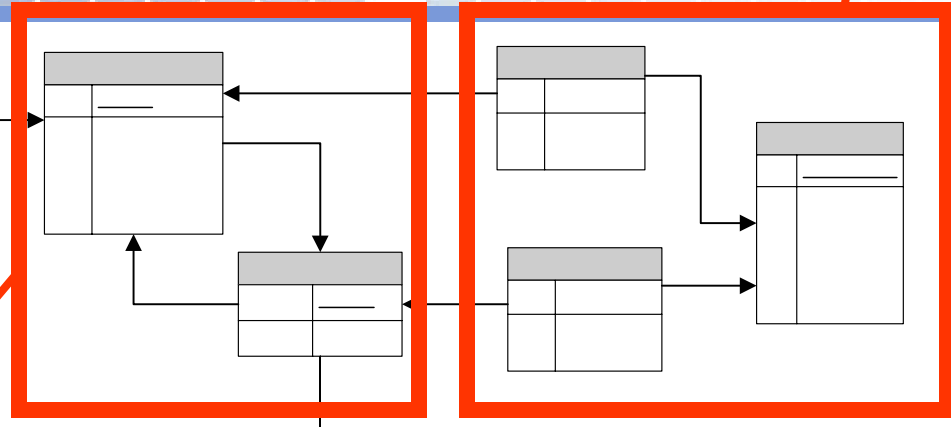
- * **Databases enable storage and manipulation of massive amounts of data**
 - * **Advanced query capabilities**
 - * **Data source integration**
 - * **Subset extraction**



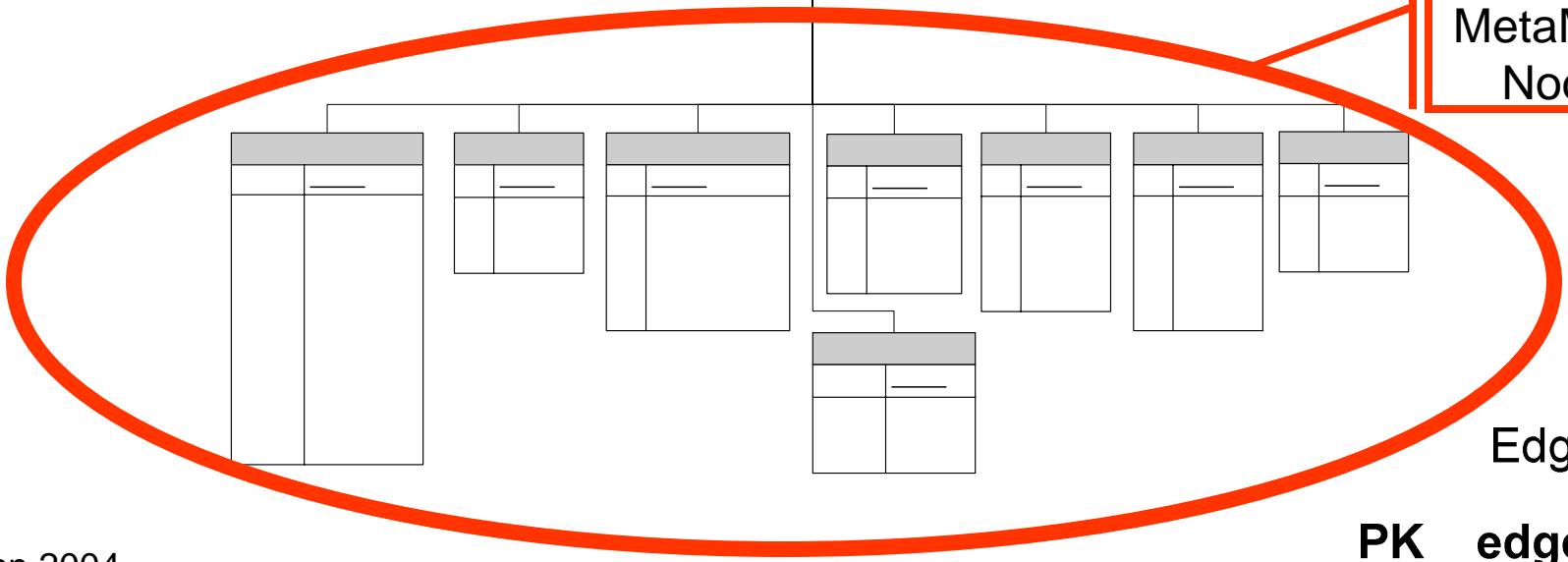
Database Schema

Data Source Tracking
IRI

Basic Graph Structure



MetaMatrix Nodes



EdgeTypes

PK edgeType

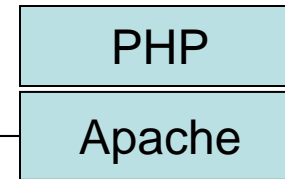


Web Clients

Nodes nearest Alluni (max distance=1)

```
SELECT * FROM casnet_128_2_72_206
```

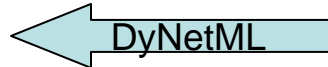
NodeID	NodeType	Edit	Export	Delete
Alluni	Agent	Edt	Exp	Del
Children	Skd	Edt	Exp	Del
Class_Record	Skd	Edt	Exp	Del
FinanceLogistics	Rule	Edt	Exp	Del
HS_Grad	Skd	Edt	Exp	Del
LT_HS	Skd	Edt	Exp	Del
Manned	Skd	Edt	Exp	Del
Name	Skd	Edt	Exp	Del
No_Children	Skd	Edt	Exp	Del
No_Manned	Skd	Edt	Exp	Del
Secode	Skd	Edt	Exp	Del
Seni-prof	Skd	Edt	Exp	Del
Some_College	Skd	Edt	Exp	Del
Thruaid	Skd	Edt	Exp	Del
Yakar	Agent	Edt	Exp	Del
Zammar	Agent	Edt	Exp	Del
	Agent	Add		



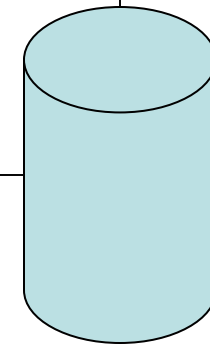
UCINET

ORA

Visualization

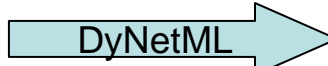


DyNetML Export



Integrated Database

AutoMap



DyNetML Import



Data Integration and Repair



- * Textual and hand-coded data often has alternative spellings or miss-spelled words
 - * Specifically, names in foreign languages
- * Database includes a *Thesaurus* table
 - * Converge alternative spellings of names to an accepted common form
 - * Transparent to the user
- * Thesauri are compatible with AutoMap



Intelligent Data Extraction



* Dataset extraction

* "Find all nodes and links referred to in a document"

* Attribute-based extraction

* "Find all American males in the dataset and extract their network"

* Ego Network extraction

* "Find all people connected to an agent"

* Graph distance-based extraction

* "Find all nodes and links within certain graph distance of a known node"

* Network Expansion

* "Given a set of nodes, find all nodes and edges in a network around them"



DyNetML Data Interchange Format



- * **Open specification**
- * **XML Derivative**
- * **Expressive**
 - * **Allows for rich social network data**
- * **Human-readable (ASCII text)**
- * **Machine-parsable with no manual intervention**
 - * **Choice of programming language/environment should not be a limiting factor**



Telling a Story with Rich Data



- * **Captures multiple interaction modes**
 - * E.g. friendship and advice networks, resource pathways
 - * Perceptual networks
- * **Captures multiple node types**
 - * E.g. people, knowledge, resources, organizations
- * **Actor attributes**
 - * Classes of nodes and edges
 - * Actor properties
 - * Allows integration with anthropological/psychological data
- * **Allows for rich network data over time**
- * **Measures and analysis results included in the same file**



DyNetML: A Rich Network Data Interchange Format



* XML-Based

- * Parsers exist in almost any programming language
 - * Less homework for implementers – no need for custom parsers
- * Open specification; human-readable

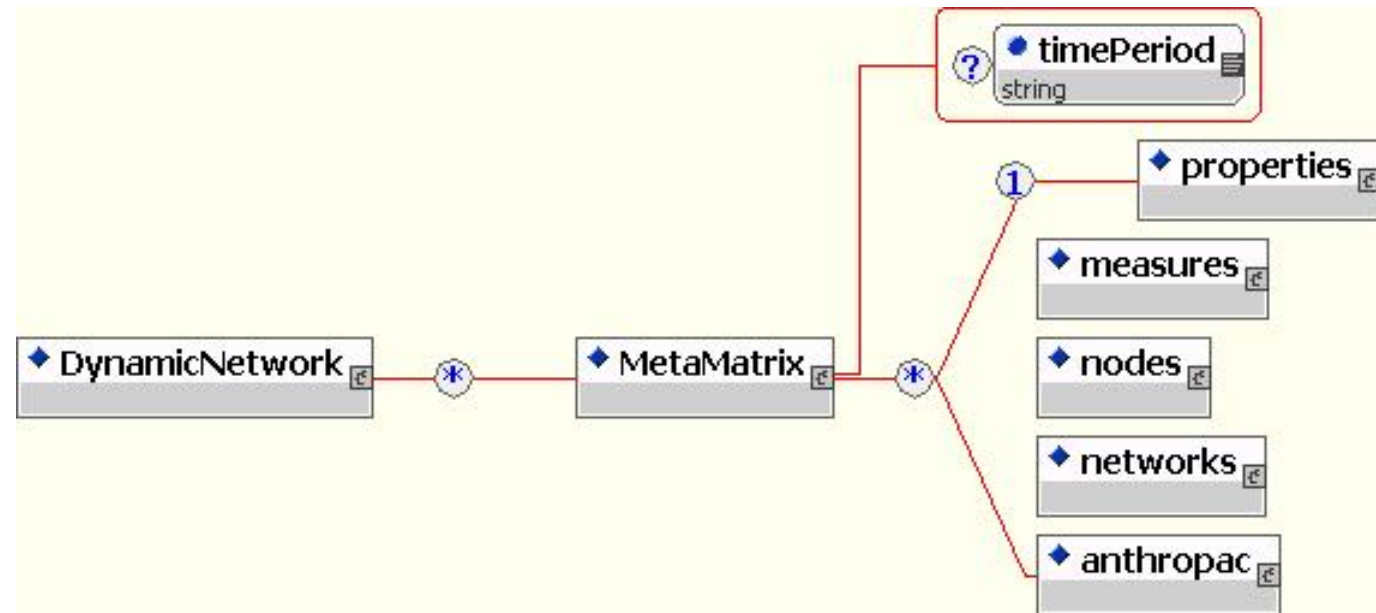
* Expressive

- * Rich network data
 - * Typed nodes and edges
 - * Extensible node and edge attributes
- * MetaMatrix specification
 - * Predefined node and matrix types can be extended
- * Rich network data over time
 - * Multiple time periods
- * Measures and analysis results included

* Extensible

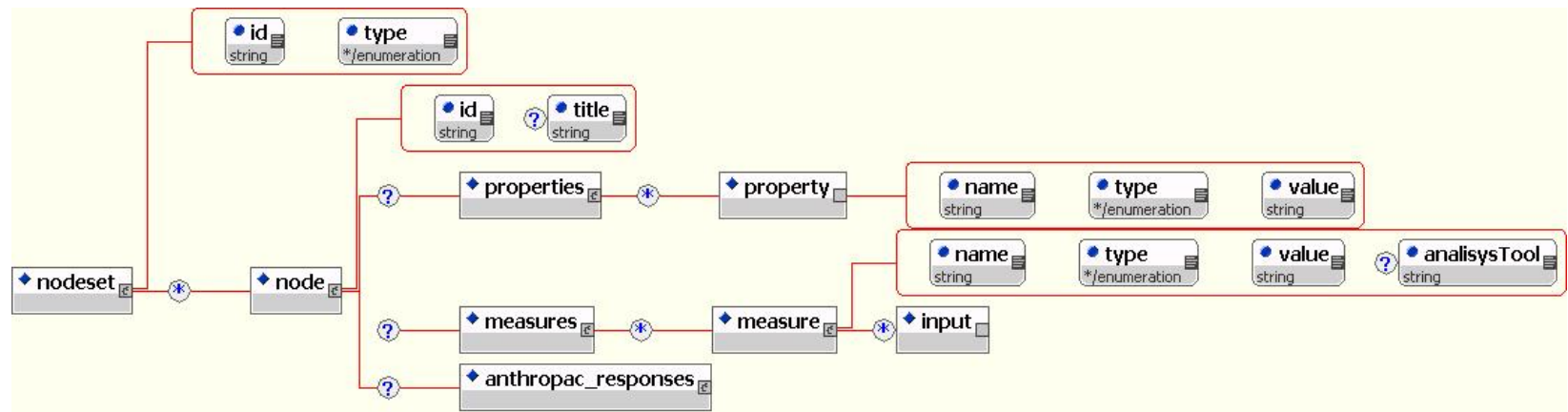
- * Adding features will not break existing code

- * A dataset is represented by a set of time periods
- * Each time period (MetaMatrix) is comprised of an arbitrary number of graphs





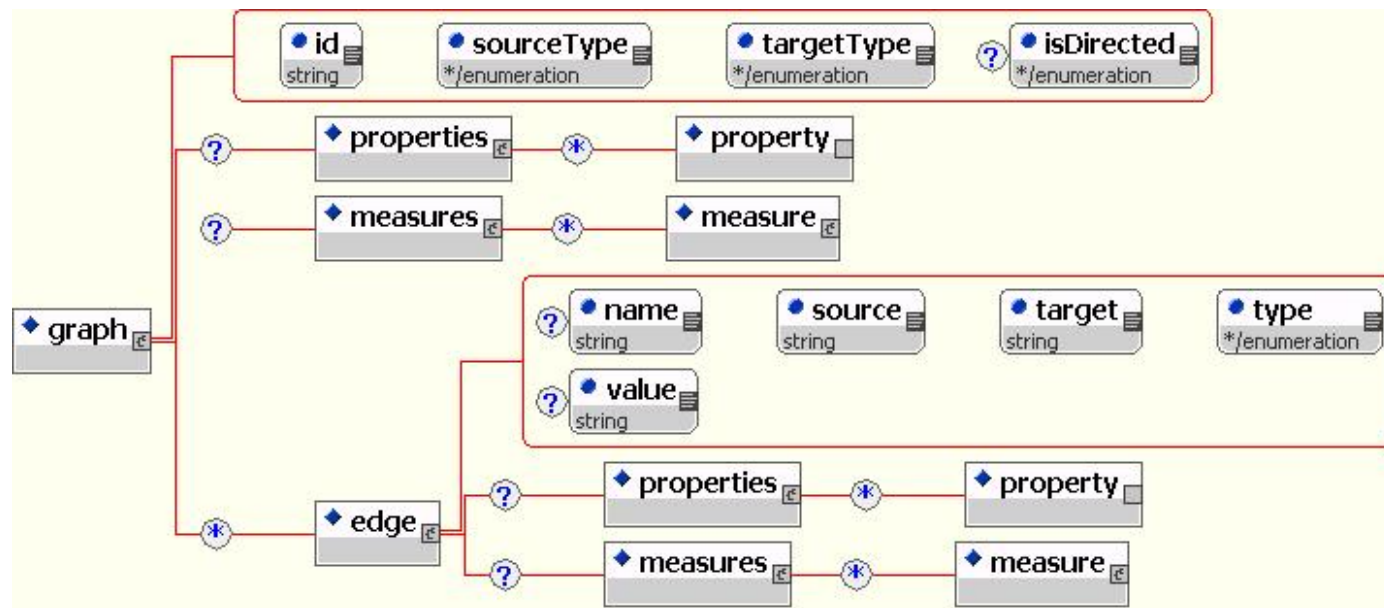
In detail... Actors and Nodes



- ✧ Nodes are grouped by type into sets
- ✧ Each node includes
 - ✧ Identifying information
 - ✧ Properties (e.g. position title, age, etc)
 - ✧ Measures – computed by analysis tools
 - ✧ Questionnaire responses



In Detail... Graphs



* Rich data for every edge

- * Captures rich fine-grained data (e.g. circumstances of data collection) and computed measures



Hands On



- * Import data derived from texts into the database
- * Explore the dataset interactively
- * Extract subsets of data for analysis
- * Analyze with Ora



Before we start



- * **Copy the DyNetML folder to your hard drive**
 - * D:\CD\CMU\DyNetML
- * **Open a command prompt window**
 - * DOS prompt
- * **Change to the DyNetML directory**
 - > `cd c:\dynetml\database`
- * **Copy the DyNetML files derived from AutoMap to c:\dynetml\database**



Dynetml2db: Export data to database



* Print out usage information

```
> dynetml2db
```

Merge information contained in a DyNetML file into the NetIntel database

```
Usage: ./dynetml2db <database name> <input file> [-h<db host> ] [-u<db username> ] [-m<message> ]
```

<database name> is the name of the database to connect to
<input file> is the name of a DyNetML input file

* This command will import a DyNetML file produced by Automap into a database named "SummerInstitute"

```
> dynetml2db SummerInstitute AutoMapResults.xml
```




Browse the Imported Data



- * Open your web browser

- * Go to

 - <http://www.casos.cs.cmu.edu/institutedb>

- * Log in

 - Username: institute

 - Password: casos2004

- * Hands on browsing with web interface



Db2dynetml: Export from Database to DyNetML



* Print out usage information

> db2dynetml

Dumps information store in a NetItel database into a DyNetML file

```
Usage: ./db2dynetml <database name> <output file> [-h<db host> ]  
[-u<db username> ] [-ego<egonet center> ] [-distance<egonet distance> ]  
[-doc<documentID> ] [-net<DyNetML file>]
```

<database name>	is the name of the database to connect to
<output file>	is the name of a DyNetML output file
-ego<egonet center>	Generate an ego network centered around a node
-distance<egonet distance>	Radius of the ego network; requires -ego
-doc<documentID>	ID of the source document
-net<DyNetML file>	grow a network from one specified in the file using consolidated data



Db2dynetml: Export from Database to DyNetML



* Export all data from database

```
> db2dynetml SummerInstitute allData.xml
```

Creates file `allData.xml` with data exported from the database



Db2dynetml: Export from Database to DyNetML



- * Export all data related to a particular document
- * In the web interface, click on “All documents”
- * Find an ID of a document you are interested in
- * **Type:**

```
> db2dynetml SummerInstitute docData.xml -doc 2  
(substitute the ID of a document you found in the web  
interface for '2')
```

Data contained in the document will be exported into a new
file named 'docData.xml'



Db2dynetml: Export from Database to DyNetML



✧ Export an ego network of a particular agent

```
> db2dynetml SummerInstitute hamas.xml -ego hamas
```

✧ Export a network of a particular agent, at a set graph distance

```
> db2dynetml SummerInstitute hamas.xml -ego hamas -  
distance 2
```

- Results can also be viewed on the web by clicking on the EGO button next to a node



Db2dynetml: Export from Database to DyNetML



* Expand a known network using data in the database

```
> db2dynetml SummerInstitute allHamas.xml -net hamas.xml
```

Searches out all nodes and links connected to any of the nodes in the hamas.xml file and extracts data about them into a new datafile



Results



* You have learned how to

- * Extract network data from texts,
- * Integrate network data from different sources using the database
- * Explore network data interactively
- * Query and export data from the database

* Proceed to running network analysis with ORA