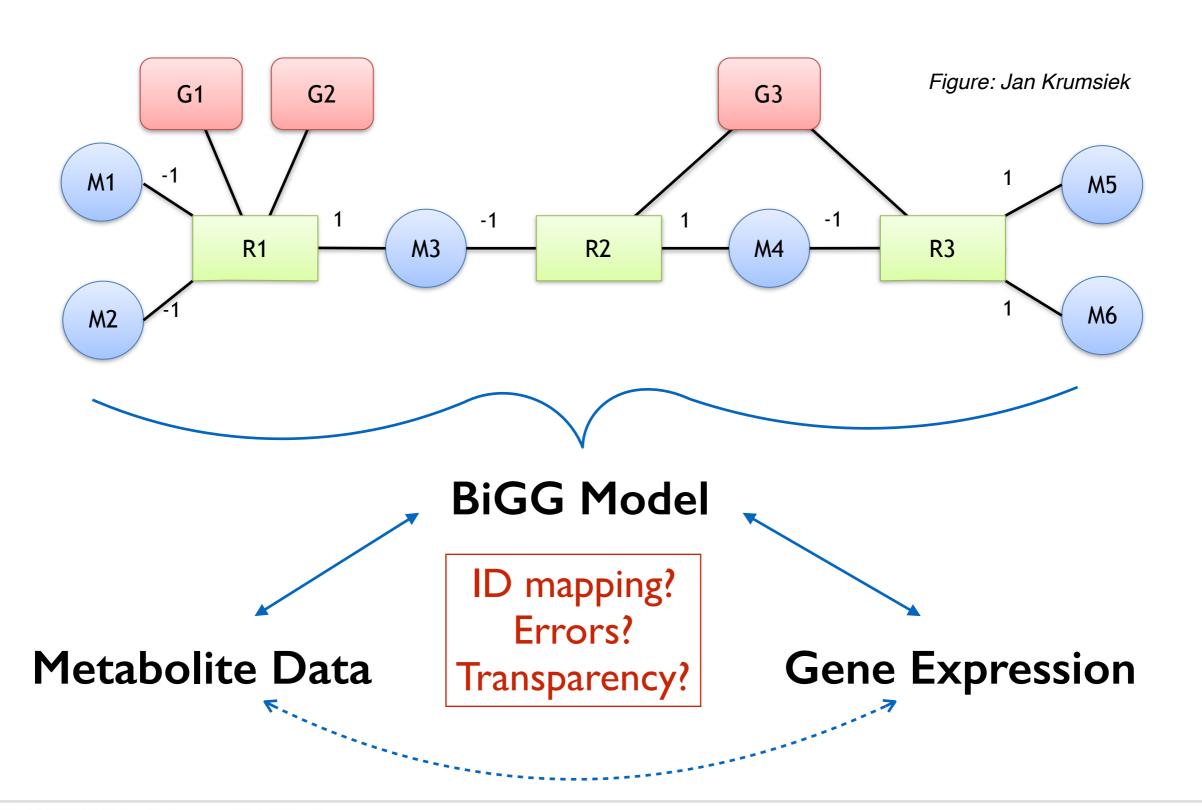
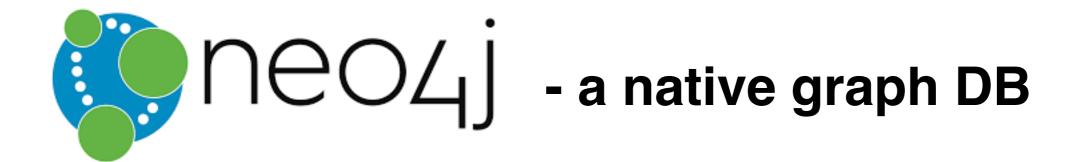
# HelmholtzZentrum münchen German Research Center for Environmental Health

# Neo4j - a graph DB

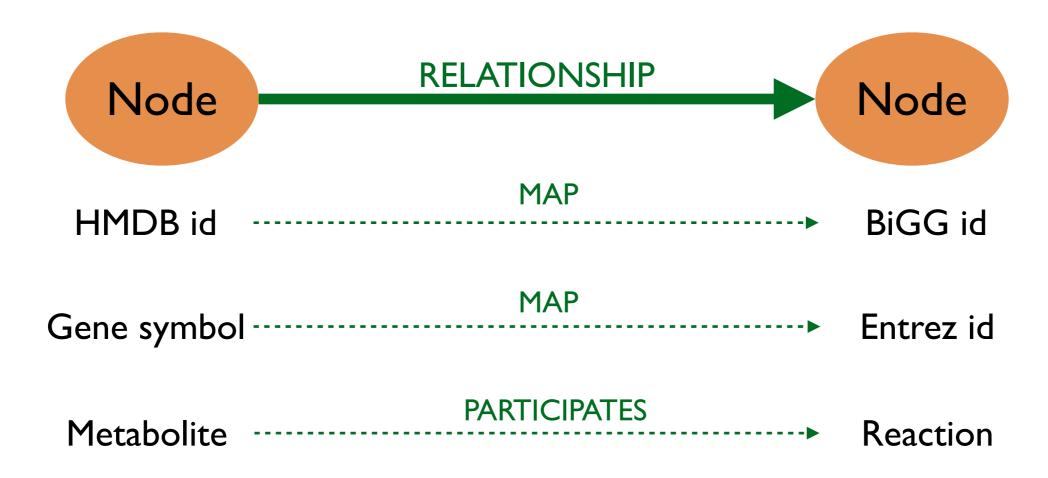
Maria Wörheide January 19, 2017

### Problematics of data integration

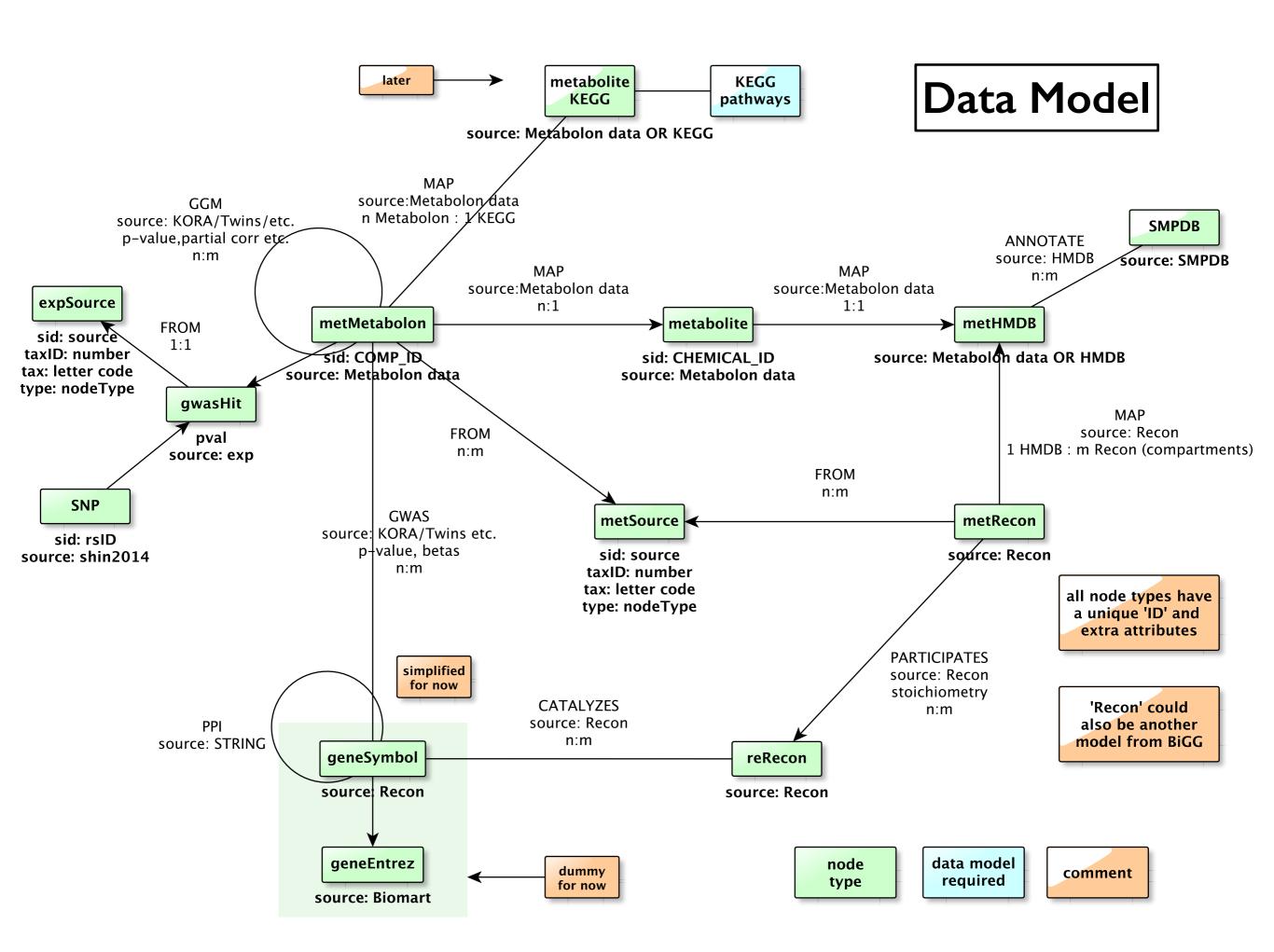




"Today's world is no longer driven by data – it's driven by the connections between them<sup>1)</sup>. "



1)https://neo4j.com/product/



Properties

### metMetabolon

sid HMDb\_ID Platform Pathway Mass

. . .

**Properties** 

metMetabolon

sid HMDb\_ID Platform Pathway Mass

• •

Label

**Properties** 

metMetabolon sid

HMDb\_ID
Platform
Pathway
Mass

• • •

Label

**Unique Property** 

**Properties** 

metMetabolon

sid
HMDb\_ID
Platform
Pathway
Mass

Label

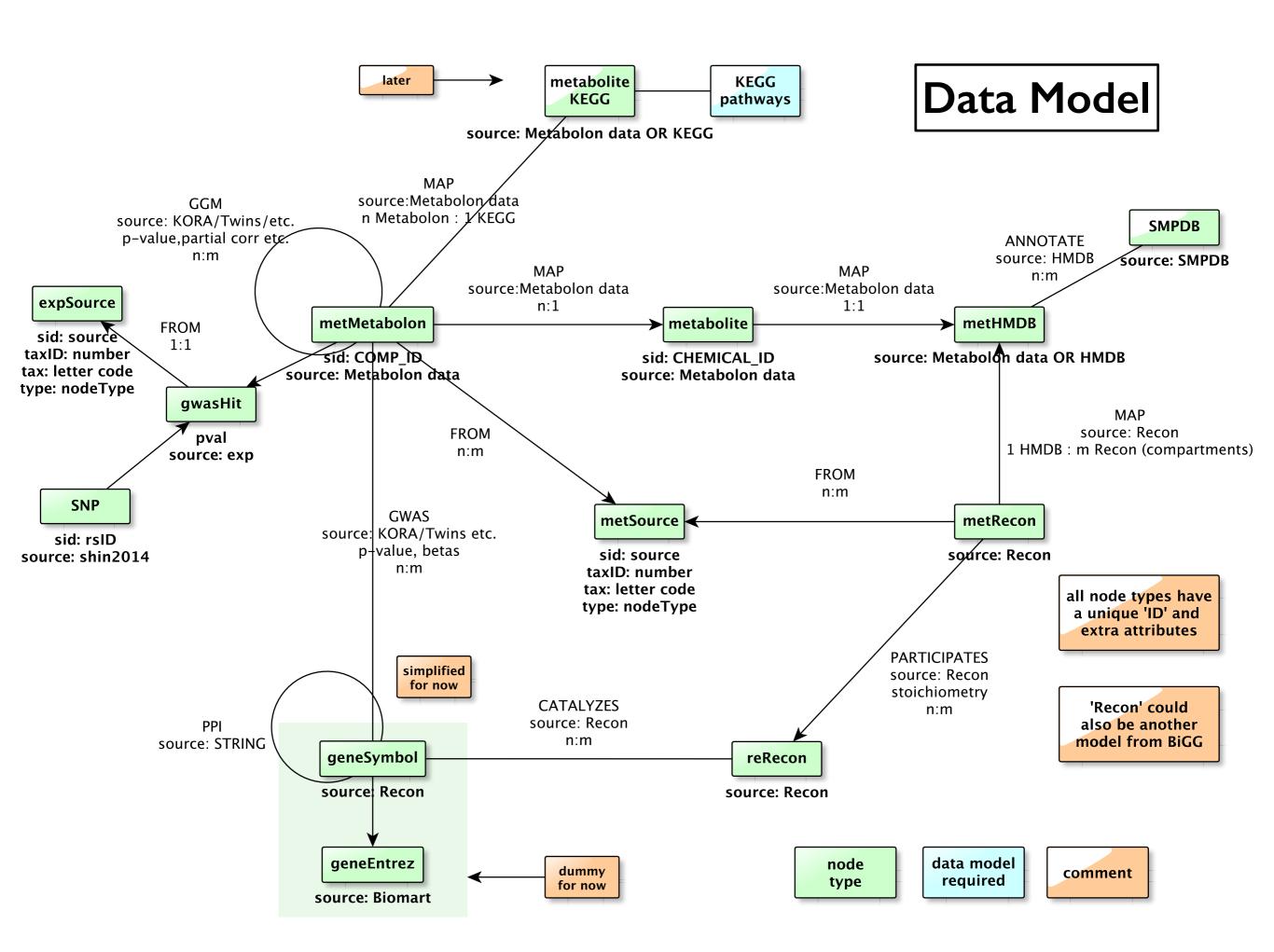
**Unique property** 

Other properties

## Neo4j - Relationships

# metMetabolon sid = 'M43264' HMDb\_ID = 'HMDBI3127' Platform = 'LC/MS Pos' Pathway = 'Lipid' Mass = '248.2' ... Properties Metabolite sid = 'I00003926' HMDb\_ID = 'HMDBI3127' BIOCHEMICAL = '3-hydroxybutyrylcarnitine (I)'

- relationships always have a direction
- → direction can be ignored in queries



## Cypher

- declarative graph query language
- uses patterns to describe graph data
- can be used for querying and updating
- queries built up using various clauses
  - familiar, SQL-like
  - can be chained together
  - intermediate results will be context for next clause

## **Queries - Simple**

Match a specific node and return it:

```
variable | label | property

MATCH (m : metabolite {sid:"1021"})
RETURN m
```

Table with all reactions that a metabolite participates in:

```
relationship type any node

MATCH (m)-[:PARTICIPATES]-(r)

RETURN m.sid AS metabo, collect(r.sid)

only value aggregation function
```

### **Queries - Advanced**

#### Return all metabolites that:

- a) participate in multiple(>1) reactions,
- b) have been measured and
- c) belong to the 'Amino acid' pathway

```
MATCH (:reRecon)-[p:PARTICIPATES]-(m:metRecon)
WITH count(p) AS nr, m AS metabo
WHERE nr>1
MATCH (metabo)-[:MAP*3]-(m:metMetabolon)
WHERE m.SUPER_PATHWAY='Amino Acid'
RETURN distinct(m.sid)
```

### **Queries - Advanced**

### Return all metabolites that:

- a) participate in multiple(>1) reactions,
- b) have been measured and
- c) belong to the 'Amino acid' pathway

```
MATCH (:reRecon)-[p:PARTICIPATES]-(m:metRecon)
WITH count(p) AS nr, m AS metabo
WHERE nr>1
MATCH (metabo)-[:MAP*3]-(m:metMetabolon)
WHERE m.SUPER_PATHWAY='Amino Acid'
RETURN distinct(m.sid)
```

### RNeo4j and Local DB

```
#install Rneo4j
install.packages("RNeo4j")
library(RNeo4j)
#connect to db
db = startGraph("http://ibisdb02:7475/db/data/")
#sample cypher query
           "MATCH (h:metHMDB)-[m:MAP]-(:metabolite)
query =
            WITH h AS hmdb, count(m) AS nr
            WHERE nr > 1
            MATCH (hmdb)-[:MAP]-(m:metabolite)
            RETURN hmdb.sid AS metabolite,collect(m.sid) AS multipleID"
#get results of query
                                  metabolite
                                                     multipleID
result <- cypher(db, query)</pre>
                                  HMDB01976
                                                    100001181, 100001580
```

#open browser view

browse(db)

... and

tick box

+ reload

page

If asked for

to browser

settings ...

**Network Connection** 

**BOLT + ROUTING** 

Do not use Bolt

URI

If unavailable Bolt will be used

password, go

# HelmholtzZentrum münchen German Research Center for Environmental Health

More information, tutorials etc.

Browser: <a href="http://ibisdb02:7475/browser/">http://ibisdb02:7475/browser/</a>

Neo4j/Graph DB:

https://neo4j.com/why-graph-databases/

Cypher:

https://neo4j.com/docs/cypher-refcard/current/

https://neo4j.com/docs/developer-manual/current/cypher/

Online tutorial:

https://neo4j.com/graphacademy/online-training/introduction-graph-databases/