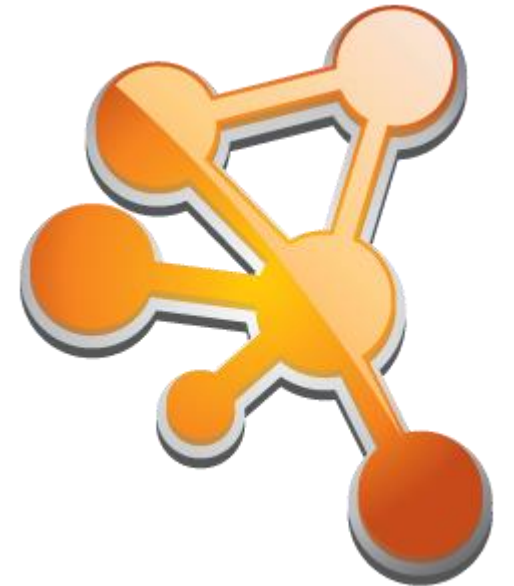# Module 8
Data Integration and Interaction Networks

10 February 2025

Dr Tomasz Kurowski
t.j.kurowski@cranfield.ac.uk

www.cranfield.ac.uk

	Mon 10/2	Tue 11/2	Wed 12/2	Thu 13/2	Fri 14/2
all-day					
06					
07					
08					
09 10	09:00 - 10:30 Introduction to Module & Data Integration B300 Ground Floor East LR	09:00 - 10:30 Database Design and Normalisation B300 Ground Floor East LR	09:00 - 10:30 Database and Web Interfaces B300 Ground Floor East LR	09:00 - 10:30 Data Integration Case Study B300 Ground Floor East LR	09:00 - 10:30 Visualising Biological Networks B300 Ground Floor East LR
11 12	11:00 - 13:00 Introduction to JavaScript B300 Ground Floor East LR	11:00 - 13:00 Database Design and Normalisation B300 Ground Floor East LR	11:00 - 13:00 Database and Web Interfaces Practical Session B300 Ground Floor East LR	11:00 - 13:00 Database and Web Interfaces Practical Session (cont.) & Course Feedback Session B300 Ground Floor East LR	11:00 - 13:00 Case Study: Single Cell Analysis B300 Ground Floor East LR
13					
14 15	14:00 - 15:30 Introduction to JavaScript Practical Session B300 Ground Floor East LR	14:00 - 15:30 Database Design and Normalisation Practical Session B300 Ground Floor East LR	14:00 - 15:30 Database and Web Interfaces Practical Session B300 Ground Floor East LR	14:00 - 15:30 Introduction to Cytoscape B300 Ground Floor East LR	14:00 - 15:30 Assignment Handling B300 Ground Floor East LR
16 17	16:00 - 17:30 Introduction to JavaScript Practical Session B300 Ground Floor East LR	16:00 - 17:30 Database Design and Normalisation Practical Session B300 Ground Floor East LR	16:00 - 17:30 Practical Catch-up Session B300 Ground Floor East LR	16:00 - 17:30 Introduction to Cytoscape Practical Session B300 Ground Floor East LR	16:00 - 17:30 Visualising Biological Networks B300 Ground Floor East LR
18					
19					

Topics covered

- Core concepts of data integration - ETL, data formats, parsing, databases, interfaces.

- JavaScript – front-end (basics) and back-end (Node, Express, API, database connectivity).

- Relational databases (normalization, SQL).

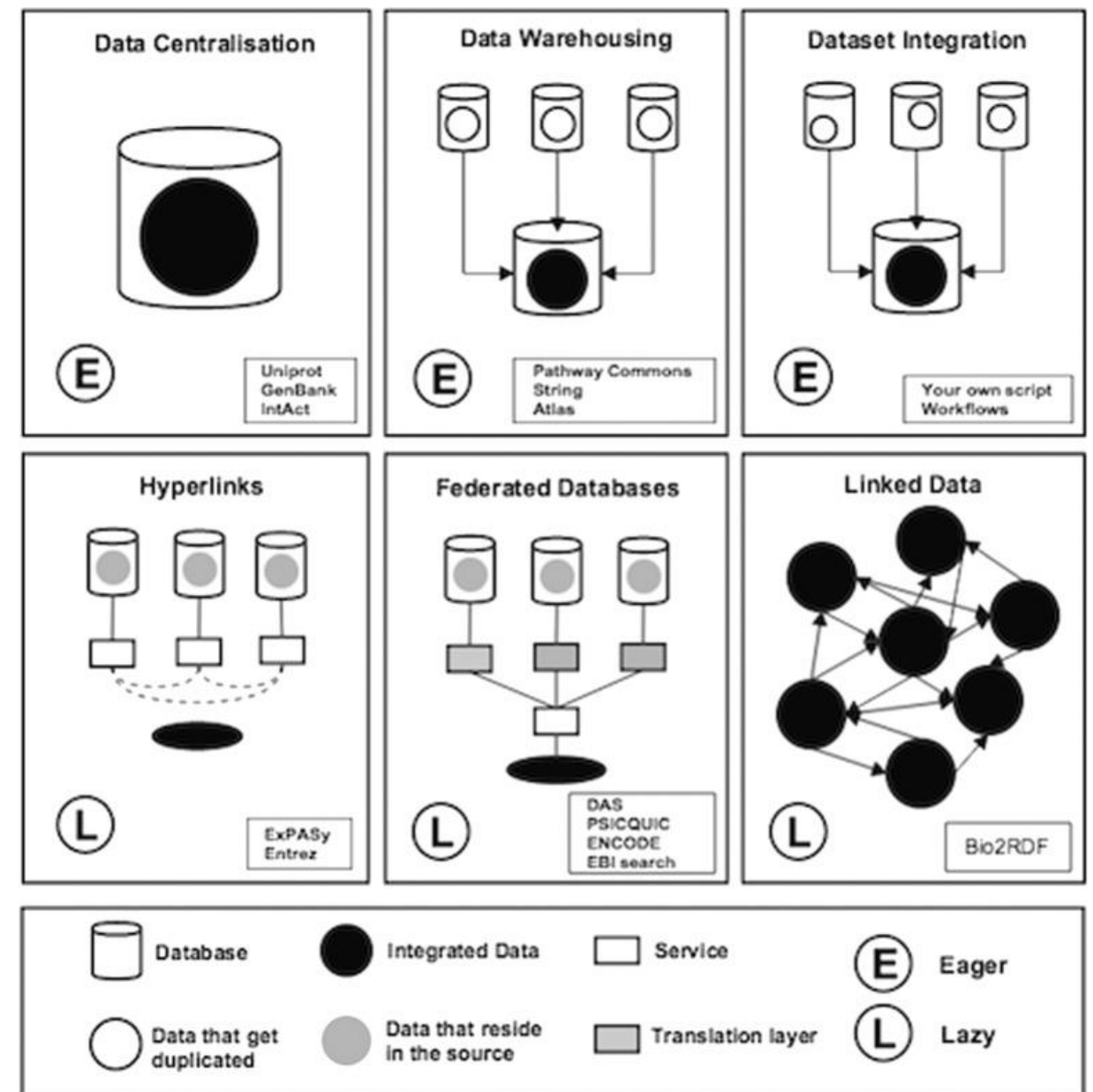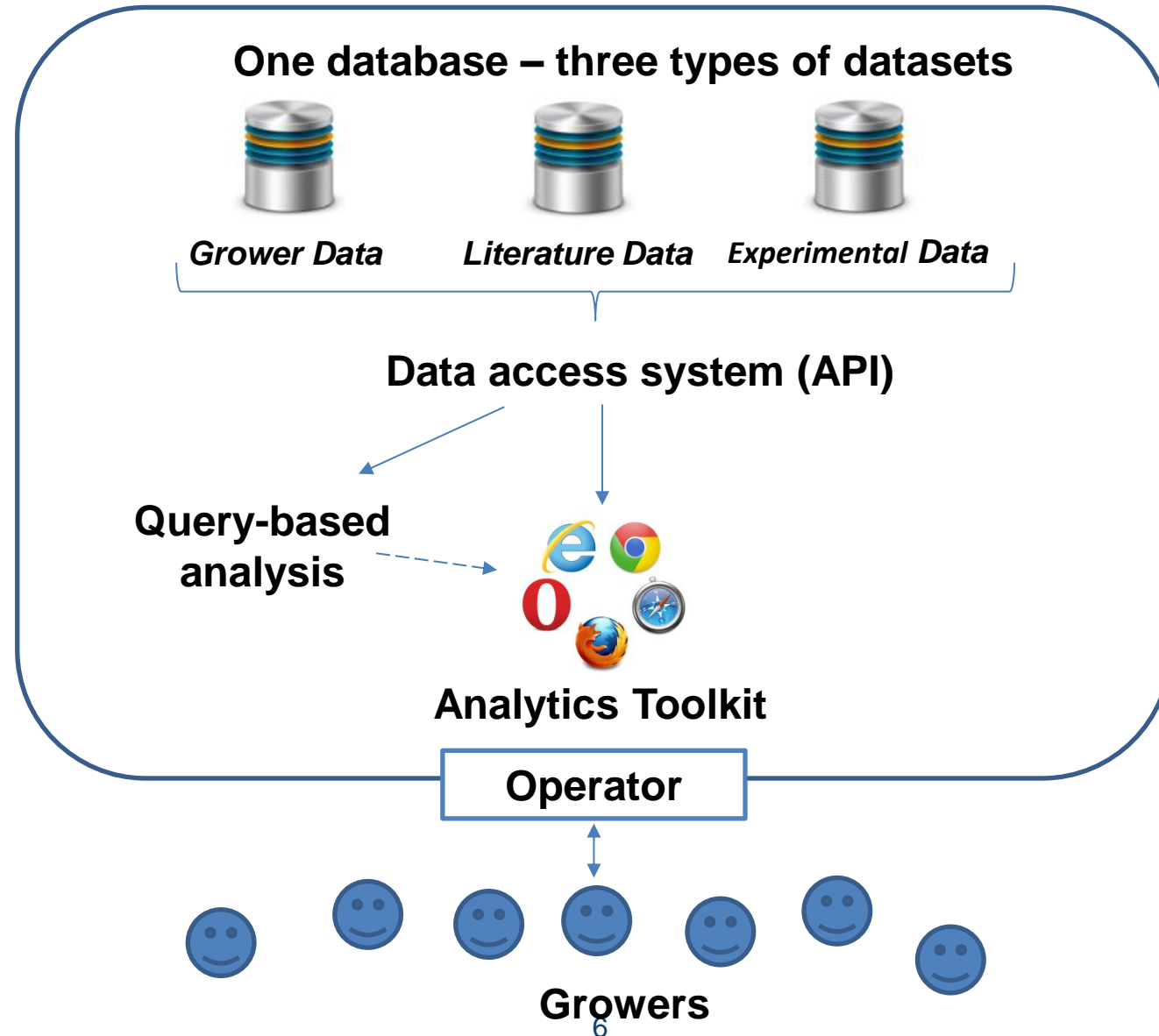- Biological interaction networks via Cytoscape, literature mining, network visualisation.
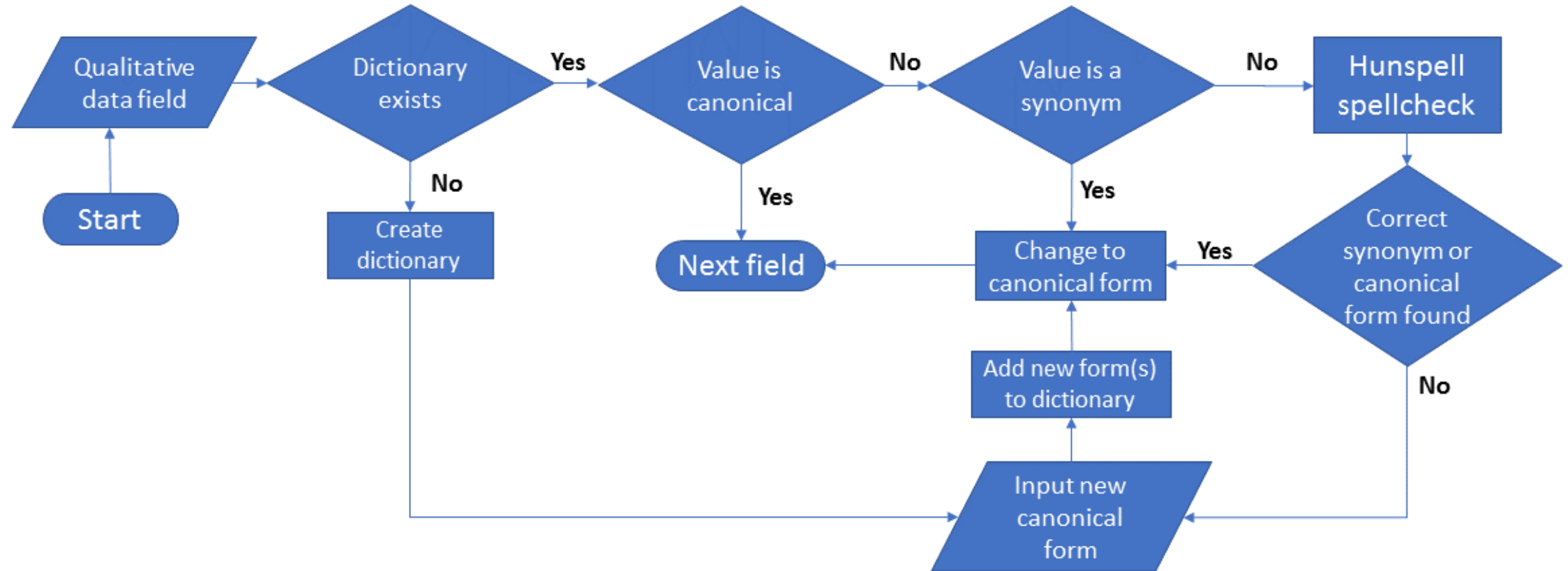
Data integration

ETL process:

- Extract – from files (CSV, JSON, XML…) or other databases

- Transform – pivoting tables, joining data from different sources, aggregation

- Load – into data warehouse for access and analytics

Lapatas, V., M. Stefanidakis, R.C. Jimenez, A. Via, and M.V. Schneider. 2015. *Data integration in biological research: an overview*. J of Biol Res-Thessaloniki 22(1). doi: 10.1186/s40709-015-0032-5.

Example: Soil Management Information System (SMIS)

One database – three types of datasets

Grower Data　　*Literature Data*　*Experimental Data*

Data access system (API)

Query-based analysis

Analytics Toolkit

Operator

Growers

"Transform"

Monday

09	09:00 - 10:30 Introduction to Module & Data Integration B300 Ground Floor East LR
10	
11	11:00 - 13:00 Introduction to JavaScript B300 Ground Floor East LR
12	
13	
14	14:00 - 15:30 Introduction to JavaScript Practical Session B300 Ground Floor East LR
15	
16	16:00 - 17:30 Introduction to JavaScript Practical Session B300 Ground Floor East LR
17	

- Module introduction (hi!).

- Storage formats, database integration, APIs, protocols, biological databases, visualisation, non-relational databases.

- Introductory JavaScript lecture and practical.

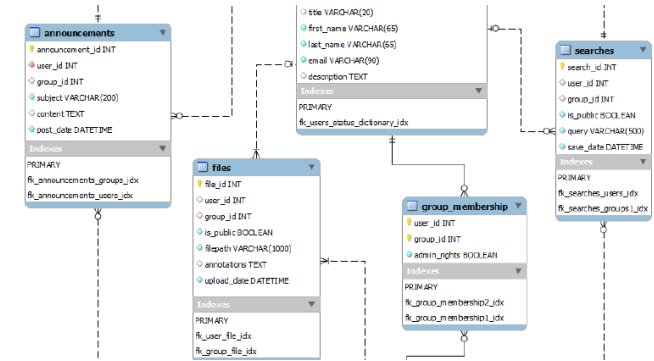- Basics of HTML / JavaScript front-end integration.

Cassandra vs **MongoDB** vs **CouchDB** vs **Redis** vs **Riak** vs **HBase** vs **Couchbase** vs **OrientDB** vs **Aerospike** vs **Neo4j** vs **Hypertable** vs **ElasticSearch** vs **Accumulo** vs **VoltDB** vs **Scalaris** vs **RethinkDB** comparison

Tuesday

09	**09:00 - 10:30** Database Design and Normalisation
10	B300 Ground Floor East LR
11	**11:00 - 13:00** Database Design and Normalisation
12	B300 Ground Floor East LR
13	
14	**14:00 - 15:30** Database Design and Normalisation Practical Session
15	B300 Ground Floor East LR
16	**16:00 - 17:30** Database Design and Normalisation Practical Session
17	B300 Ground Floor East LR

- Relational model, relational database management systems (RDBMS), structured query language (SQL).

- Database design and normalisation.

- Database connectivity.

Wednesday

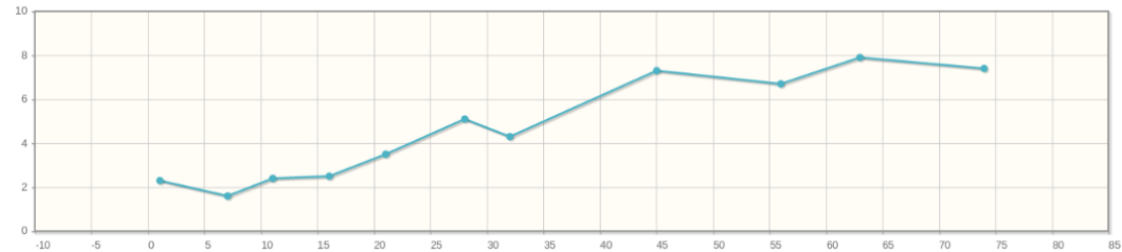|	
09	**09:00 - 10:30** Database and Web Interfaces B300 Ground Floor East LR
10	
11	**11:00 - 13:00** Database and Web Interfaces Practical Session B300 Ground Floor East LR
12	
13	
14	**14:00 - 15:30** Database and Web Interfaces Practical Session B300 Ground Floor East LR
15	
16	**16:00 - 17:30** Practical Catch-up Session B300 Ground Floor East LR
17	

- Database interfaces with JavaScript.

- REST API implementation using Express.

- (Catch-up session and optional front-end practical)

Thursday

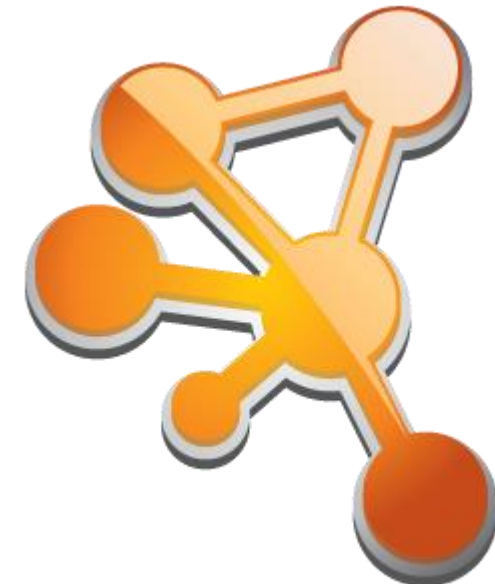| 09 | 09:00 - 10:30
Data Integration Case Study
B300 Ground Floor East LR |
10	
11	11:00 - 13:00 Database and Web Interfaces Practical Session (cont.) & Course Feedback Session B300 Ground Floor East LR
12	
13	
14	14:00 - 15:30 Introduction to Cytoscape B300 Ground Floor East LR
15	
16	16:00 - 17:30 Introduction to Cytoscape Practical Session B300 Ground Floor East LR
17	

- Data integration case studies (VarGen, Tersect Browser).

- Course feedback session with SAS at ~12:00!

- Continuation of JavaScript REST API work (including basic visualisation components).

- Introduction to Cytoscape and biological networks.

Friday
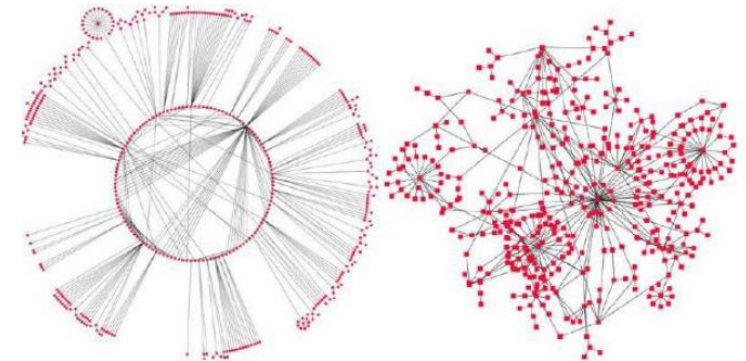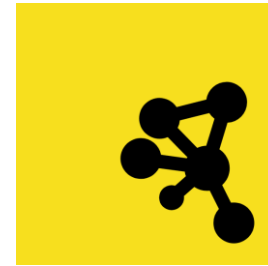
09	**09:00 - 10:30** Visualising Biological Networks B300 Ground Floor East LR
10	
11	**11:00 - 13:00** Case Study: Single Cell Analysis B300 Ground Floor East LR
12	
13	
14	**14:00 - 15:30** Assignment Handling B300 Ground Floor East LR
15	
16	**16:00 - 17:30** Visualising Biological Networks B300 Ground Floor East LR
17	

- Visualising biological networks in Cytoscape.

- Case Study: Single Cell Analysis with Dr Krzysztof Polanski (Cambridge Stem Cell Institute)

- Generating networks based on literature mining.

- Assignment!

- (Optional OSGi practical – Cytoscape Apps).

Cambridge Stem Cell Institute

Cranfield University

Module 8 Assignment
Data Integration and
Interaction Networks

Dr Tomasz Kurowski
t.j.kurowski@cranfield.ac.uk

OSGi Alliance

Module engagement QR code

If you are unable to scan this code, please contact SAS Admin – **seeaadmin@cranfield.ac.uk**