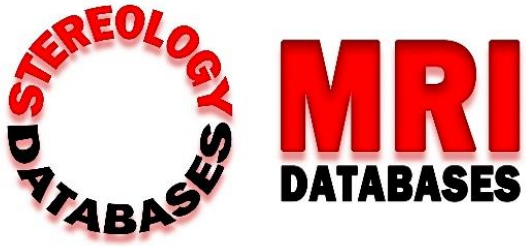
INTRODUCTION TO COMPLEXITY USING RELATIONAL DATABASES

Enterprise Biology Software Project - 2016



Biology operates under the rules of complexity, wherein it combines parts and connections in ways that produce desirable outcomes. In contrast, we prefer to access just biological parts, but not their connections. This puts us at a considerable disadvantage because it prevents us from exploring biology as a complexity. In consequence, our research runs the risk of being incompatible with biology. The databases described herein were designed specifically to address this issue in a concrete way.

We now know that research data stored in relational databases can be used to recover the connections and complexity previously discarded by our experimental methods. By recovering the missing information, the rule-based patterns of biology can appear and spark a host of novel solutions (Bolender 2016). Often, the results can be startling. When a database reaches its critical mass, complexity kicks in and we can begin to see what's happening in biology's world. We also discover what we've been missing.

Patterns are the currency of complexity. We can use them to detect changes by comparing one pattern to another or by noting the presence or absence thereof. Solving complex problems with patterns, however, requires a sequence of at least three types of databases. The first of these houses the data taken from the biology literature, the second one translates the data into complex patterns, and the third one tunes the patterns to solve specific problems. In short, the databases transform the biology literature into parallel complexities capable of emulating biology. Such an unusual approach to complexity creates an even more unusual relationship with biology. We submit problems to biology quantitatively, and then let biology deliver solutions consistent with its set of rules. This results is a collaboration, wherein we learn how to ask questions that we think only biology can answer. Representative examples of these three types of databases can be found below.

1. Parts Catalogue

BIOLOGY LITERATURE – SIMPLE DATA (PARTS)		
Original Data →	**Stereology**	**MRI**
Data Types	V, S, L, N	V
Species	Multiple (Human, etc.)	Single (Human)
Subjects	Postmortem	Living
Data Sources	Reprints, Literature, Internet	IBVD

The parts catalogue includes data transferred from referred publications and stored in relational databases. Such data include the numerical data of named parts.

2. Patterns Catalogue

DATABASES (PATTERNS) – COMPLEX DATA (PARTS & CONNECTIONS)		
Derived Data →	**Stereology**	**MRI**
Data Types	V, S, L, N	V
Species	Multiple (Human, etc.)	Single (Human)
Subjects	Postmortem	Living
Data Sources	Reprints, Literature	IBVD, Literature
Mathematical Markers		
• Data Pairs	X	
• Data Triplets	X	X
• Data Quadruplets		X
Connection Ratios		
• Triplets	X	X

The relationship of one part to another can be expressed quantitatively as a ratio. This ratio recovers the connectivity of the parts. Taken together, parts and connections define biological complexity with patterns. These patterns translate into units of complexity, which can be represented quantitatively as mathematical markers. A triplet marker (AX:BY:CZ), for example, includes three named parts (A:B:C) with a corresponding ratio (X:Y:Z). Such markers, which can be generated from the literature in large numbers, serve as a direct link to biological complexity. Moreover, the markers provide open access to the mathematical principles of biology, which play a key role in solving complex problems. The understanding to come from this reformatting of the biology literature is that relational databases can allow us to explore biology as a quantitative science.

3. Patterns as Solutions

DATABASES (TUNED) – COMPLEX DATA (PARTS & CONNECTIONS)								
Solutions	**MM**	**CR**	**Pairs**	**Triplets**	**Quads**	**Stereology**	**MRI**	**Report**
Disorders of the Brain	X			X	X		X	2012
Volume Dependent Data	X		X	X		X	X	2013
Diagnosis (Data Cage)	X			X	X		X	2014
Patterns of Brain Disorders	X			X			X	2015
Reproducibility Test	X	X		X		X	X	2016

*As described in **Playing the Complexity Game with Biology**, we can extend our research options - quite remarkably - by operating under the aegis of complexity theory. The game consists of generating large sets of patterns and then selecting certain combinations thereof to solve specific problems. For example, studying the disease process in the human brain benefits importantly from a complete set of patterns, whereas making a clinical diagnosis prefers just unique patterns and running a reproducibility test just duplicate patterns. The game is an interesting one to play because we can enter – albeit it momentarily – into the world of a very smart biology.*

ONLINE DATABASES – STEREOLOGY – POSTMORTEM DATA						
DB Type	**Solution**	**Source**	**Data Type**	**Database File Name**	**Rows**	**Report**
1	Biology Literature	Stereology	V, S, L, N	1_stereology_catalogue_2016	1000	2001
2	Patterns – Co/Ex	Stereology	V, S, N	2_stereology_triplets_2016	1000	2012-15
3	Reproducibility Test	Stereology	N	3_repro_test_t_n_2016	1000	2016
3	Reproducibility Test	Stereology	S	4_repro_test_t_s_2016	1000	2016

Database (DB) Type 1 – Original data as published
Database (DB) Type 2 – Patterns defined by mathematical markers (MM)
Database (DB) Type 3 – Patterns selected to solve a given problem (e.g., connection ratios (CR))

ONLINE DATABASES – MRI – PATIENT DATA						
DB Type	**Solution**	**Source**	**Data Type**	**File Name**	**Rows**	**Report**
1	Biology Literature	MRI	V	Online (IBVD)		2012
2	Patterns - Health/ Disease	MRI	V	5_mri_triplets_2016	1000	2012-13
3	Diagnosis – Data Cage	MRI	V	6_mri_t_diagnosis_2016	1000	2014
3	Diagnosis – Data Cage	MRI	V	7_mri_q_diagnosis_2016	1000	2014
3	Patterns - Generalize	MRI	V	8_mri_t_dups_2016	1000	2015
3	Reproducibility Test	MRI	V	9_repro_test_t_v_2016	1000	2016

The complete database files can be downloaded as text files (*.cvs) that can be imported into Excel or Access.

DOWNLOAD DATABASES (*.ZIP) – STEREOLOGY – POSTMORTEM DATA						
DB Type	**Solution**	**Source**	**Data Type**	**Database File Name**	**Rows**	**Report**
1	Biology Literature	Stereology	V, S, L, N	1_stereology_catalogue_2016	24,769	2001
2	Patterns – Co/Ex	Stereology	V, S, N	2_stereology_triplets_2016	127,328	2012-15
3	Reproducibility Test	Stereology	N	3_repro_test_t_n_2016	1,236	2016
3	Reproducibility Test	Stereology	S	4_repro_test_t_s_2016	21,306	2016

DOWNLOAD DATABASES (*.ZIP) – MRI – PATIENT DATA						
DB Type	**Solution**	**Source**	**Data Type**	**File Name**	**Rows**	**Report**
1	Biology Literature	MRI	V	Online (IBVD)		2012
2	Patterns - Health/ Disease	MRI	V	5_mri_triplets_2016	723,508	2012-13
3	Diagnosis – Data Cage	MRI	V	6_mri_t_diagnosis_2016	277,039	2014
3	Diagnosis – Data Cage	MRI	V	7_mri_q_diagnosis_2016	1,048,575	2014
3	Patterns - Generalize	MRI	V	8_mri_t_dups_2016	360,956	2015
3	Reproducibility Test	MRI	V	9_repro_test_t_v_2016	379,906	2016