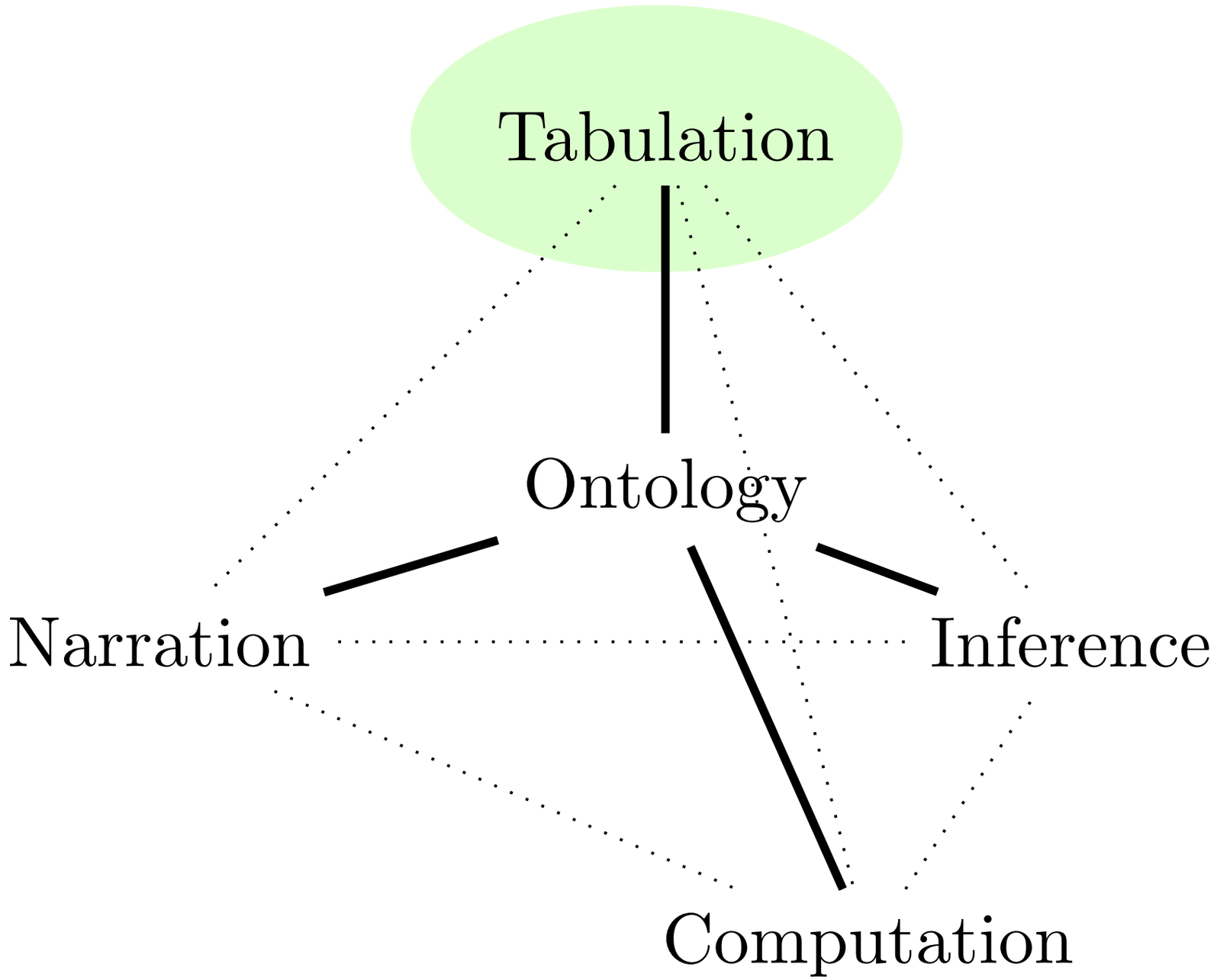
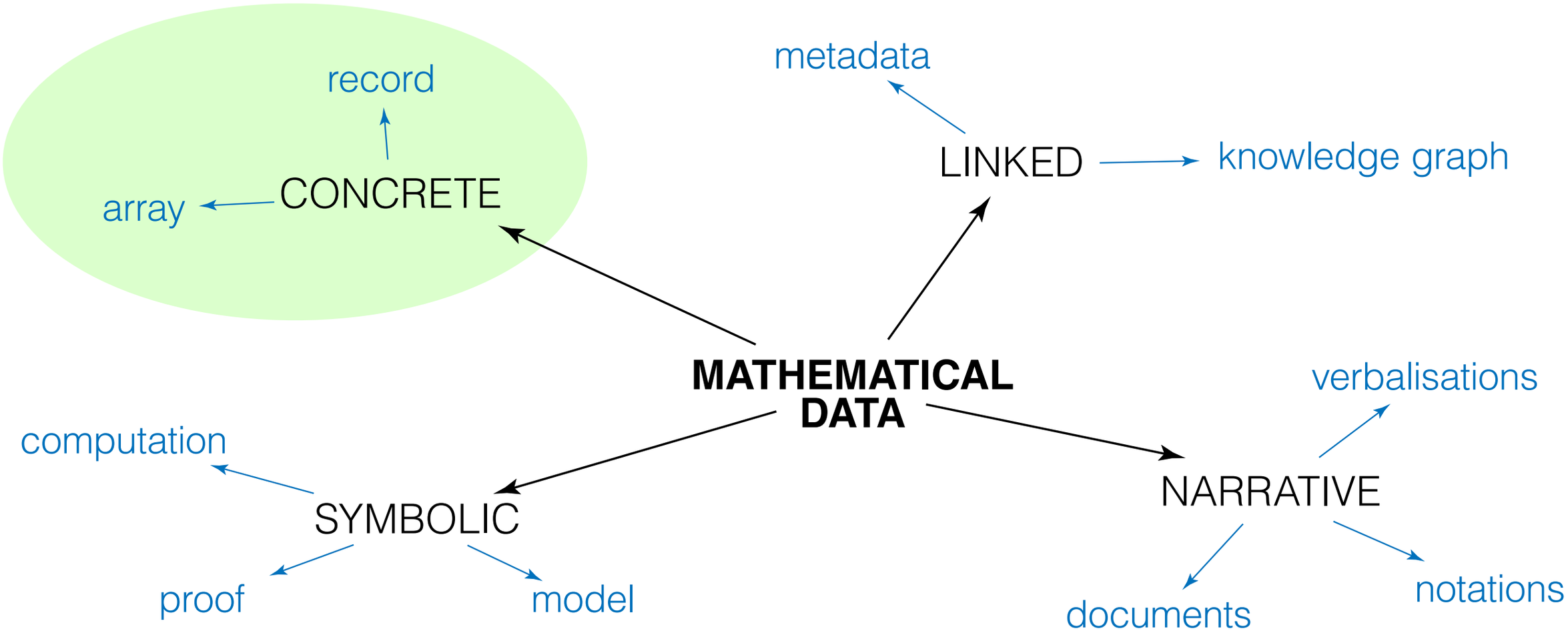
Towards a unified mathematical data infrastructure

Katja Berčič
Michael Kohlhase
Florian Rabe

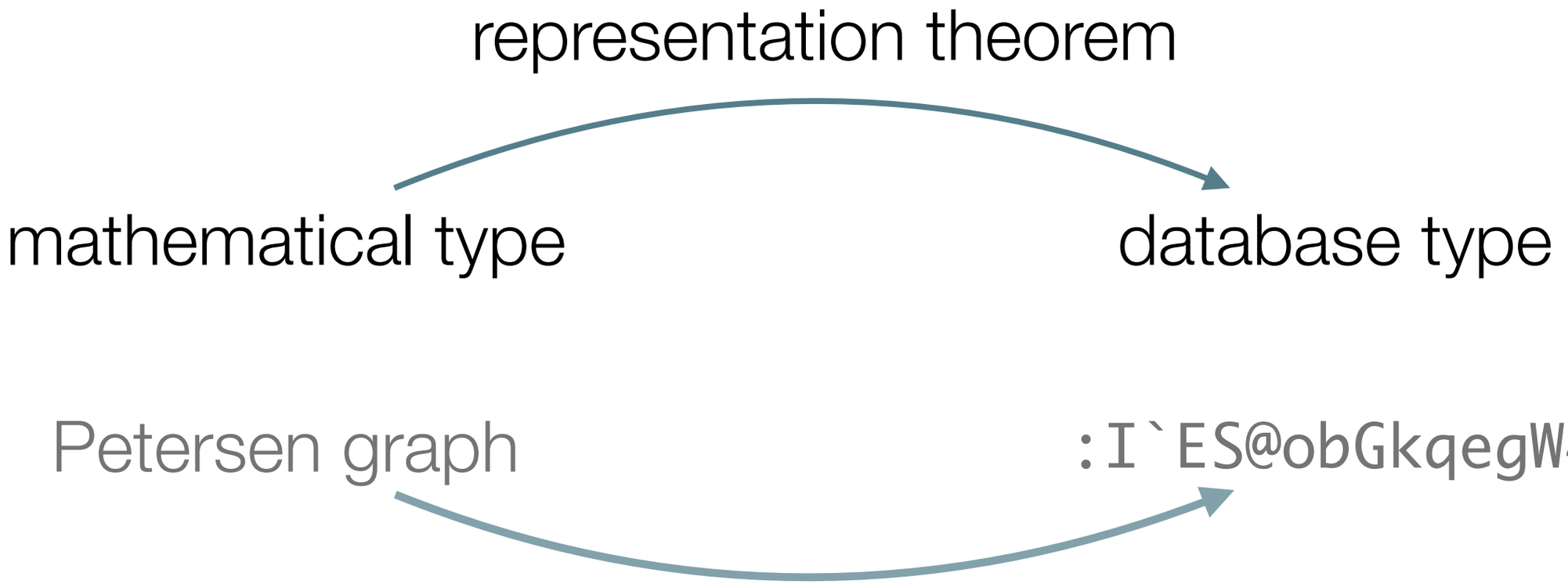
The Tetrapod



Zoom: Tabulation



Concrete data



An example

[N,i]	V	E	Tr	W?	B?	|AG|	vs	ds	#STO	gi
C4[5,1]	5	10	DT	W	NB	120	24	6	0	3
C4[6,1]	6	12	DT	U	NB	48	8	2	1	3
C4[8,1]	8	16	DT	U	Bip	(2^7)(3^2)	144	36	2	4
C4[9,1]	9	18	DT	W	NB	72	8	2	1	3
C4[10,1]	10	20	DT	U	NB	320	32	8	1	4
C4[10,2]	10	20	DT	W	Bip	240	24	6	1	4
C4[12,1]	12	24	DT	U	Bip	768	64	16	3	4
C4[12,2]	12	24	DT	W	NB	48	4	1	2	3
C4[13,1]	13	26	DT	W	NB	52	4	1	0	4
C4[14,1]	14	28	DT	U	NB	(2^8)(7^1)	128	32	1	4
C4[14,2]	14	28	DT	W	Bip	336	24	6	0	4
C4[15,1]	15	30	DT	W	NB	60	4	1	2	4
C4[15,2]	15	30	DT	W	NB	120	8	2	0	3
C4[16,1]	16	32	DT	U	Bip	(2^12)	256	64	3	4
C4[16,2]	16	32	DT	W	Bip	384	24	6	2	4

Wilson, Potočnik; A Census of edge-transitive tetravalent graphs

Summary (revisited)

We want to make

a certain kind of datasets in mathematics

more useful

by building a hosting platform.

[previous slide]
more common than one would expect

easier to share, searchable, FAIR (**f**indable, **a**ccessible, **i**nteroperable, **r**eusable - big in research data)

best bang for the buck
[see the rest of the talk]

History: before the systematic use of computers

TABLE 1
Logarithms of Numbers

1000–1500

No.	0	d	1	d	2	d	3	d	4	d	5	d	6	d	7	d	8	d	9	d
100	00000	43	00043	44	00087	43	00130	43	00173	44	00217	43	00260	43	00303	43	00346	43	00389	43
101	00432	43	00475	43	00518	43	00561	43	00604	43	00647	42	00689	43	00732	43	00775	42	00817	43
102	00860	43	00903	42	00945	43	00988	42	01030	42	01072	43	01115	42	01157	42	01199	43	01242	42
103	01284	42	01326	42	01368	42	01410	42	01452	42	01494	42	01536	42	01578	42	01620	42	01662	41
104	01703	42	01745	42	01787	41	01828	42	01870	42	01912	41	01953	42	01995	41	02036	42	02078	41
105	02119	41	02160	42	02202	41	02243	41	02284	41	02325	41	02366	41	02407	42	02449	41	02490	41
106	02531	41	02572	40	02612	41	02653	41	02694	41	02735	41	02776	40	02816	41	02857	41	02898	40
107	02938	41	02979	40	03019	41	03060	40	03100	41	03141	40	03181	41	03222	40	03262	40	03302	40
108	03342	41	03383	40	03423	40	03463	40	03503	40	03543	40	03583	40	03623	40	03663	40	03703	40
109	03743	39	03782	40	03822	40	03862	40	03902	39	03941	40	03981	40	04021	39	04060	40	04100	39
110	04139	40	04179	39	04218	40	04258	39	04297	39	04336	40	04376	39	04415	39	04454	39	04493	39
111	04532	39	04571	39	04610	40	04650	39	04689	38	04727	39	04766	39	04805	39	04844	39	04883	39
112	04922	39	04961	38	04999	39	05038	39	05077	38	05115	39	05154	38	05192	39	05231	38	05269	39
113	05308	38	05346	39	05385	38	05423	38	05461	39	05500	38	05538	38	05576	38	05614	38	05652	38
114	05690	39	05729	38	05767	38	05805	38	05843	38	05881	37	05918	38	05956	38	05994	38	06032	38
115	06070	38	06108	37	06145	38	06183	38	06221	37	06258	38	06296	37	06333	38	06371	37	06408	38
116	06446	37	06483	38	06521	37	06558	37	06595	38	06633	37	06670	37	06707	37	06744	37	06781	38
117	06819	37	06856	37	06893	37	06930	37	06967	37	07004	37	07041	37	07078	37	07115	36	07151	37
118	07188	37	07225	37	07262	36	07298	37	07335	37	07372	36	07408	37	07445	37	07482	36	07518	37
119	07555	36	07591	37	07628	36	07664	36	07700	37	07737	36	07773	36	07809	37	07846	36	07882	36
120	07918	36	07954	36	07990	37	08027	36	08063	36	08099	36	08135	36	08171	36	08207	36	08243	36
121	08279	35	08314	36	08350	36	08386	36	08422	36	08458	35	08493	36	08529	36	08565	35	08600	36
122	08636	36	08672	35	08707	36	08743	35	08778	36	08814	35	08849	35	08884	36	08920	35	08955	36
123	08991	35	09026	35	09061	35	09096	36	09132	35	09167	35	09202	35	09237	35	09272	35	09307	35
124	09342	35	09377	35	09412	35	09447	35	09482	35	09517	35	09552	35	09587	34	09621	35	09656	35
125	09691	35	09726	34	09760	35	09795	35	09830	34	09864	35	09899	35	09934	34	09968	35	10003	34
126	10037	35	10072	34	10106	34	10140	35	10175	34	10209	34	10243	35	10278	34	10312	34	10346	34
127	10380	35	10415	34	10449	34	10483	34	10517	34	10551	34	10585	34	10619	34	10653	34	10687	34
128	10721	34	10755	34	10789	34	10823	34	10857	33	10890	34	10924	34	10958	34	10992	33	11025	34
129	11059	34	11093	33	11126	34	11160	33	11193	34	11227	34	11261	33	11294	33	11327	34	11361	33
130	11394	34	11428	33	11461	33	11494	34	11528	33	11561	33	11594	34	11628	33	11661	33	11694	33
131	11727	33	11760	33	11793	33	11826	34	11860	33	11893	33	11926	33	11959	33	11992	32	12024	33
132	12057	33	12090	33	12123	33	12156	33	12189	33	12222	32	12254	33	12287	33	12320	32	12352	33

logarithm tables (17th century)

The Foster census of cubic symmetric graphs (begun in 1930, book published in 1988)


THE FOSTER CENSUS

R.M. FOSTER'S CENSUS

OF

CONNECTED SYMMETRIC TRIVALENT GRAPHS

extended and edited

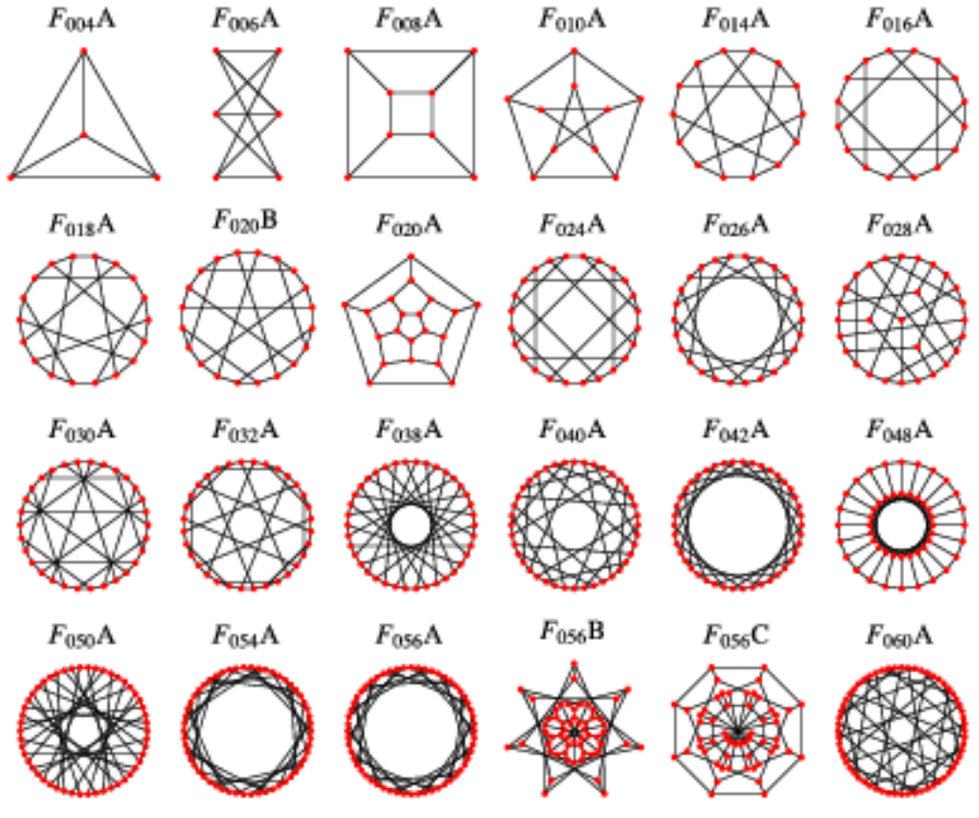
by

I.Z. BOUWER

Co-editors: W.W. Chernoff, B. Monson and Z. Star

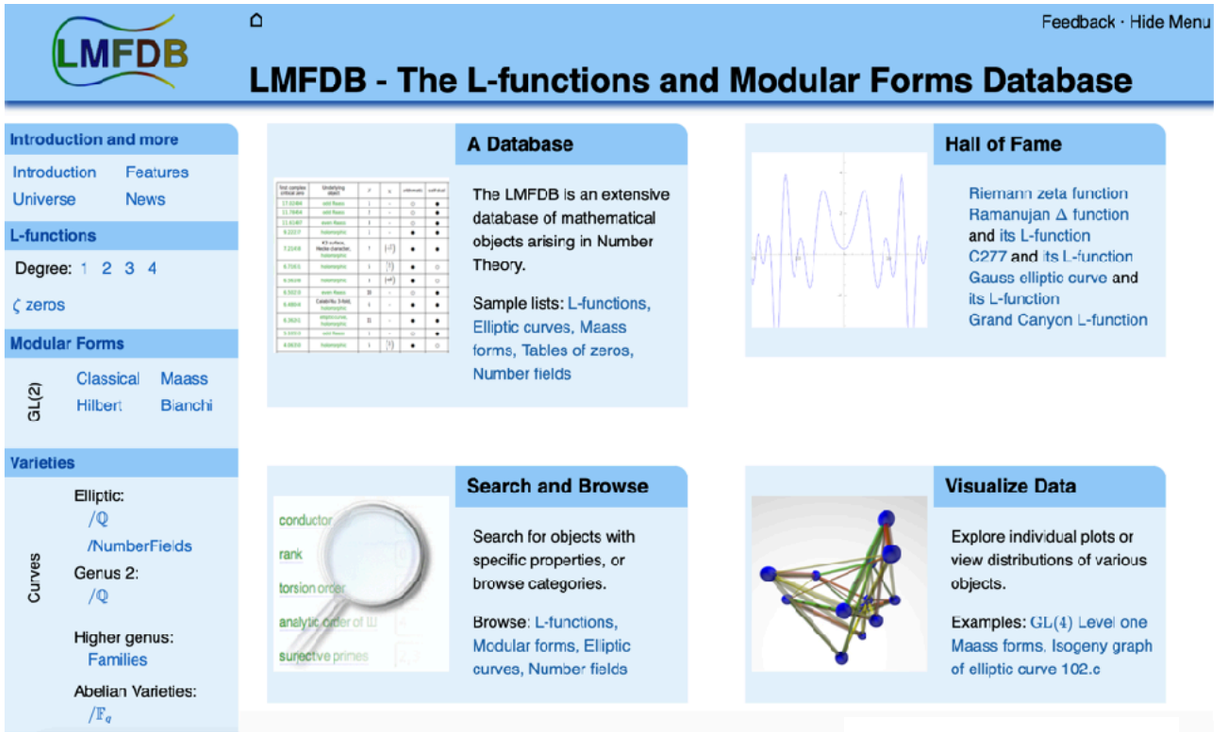
Department of Mathematics
University of New Brunswick
Fredericton, New Brunswick
Canada

With a Foreword by H.S.M. Coxeter
and a Biographical Preface by L. Schuster

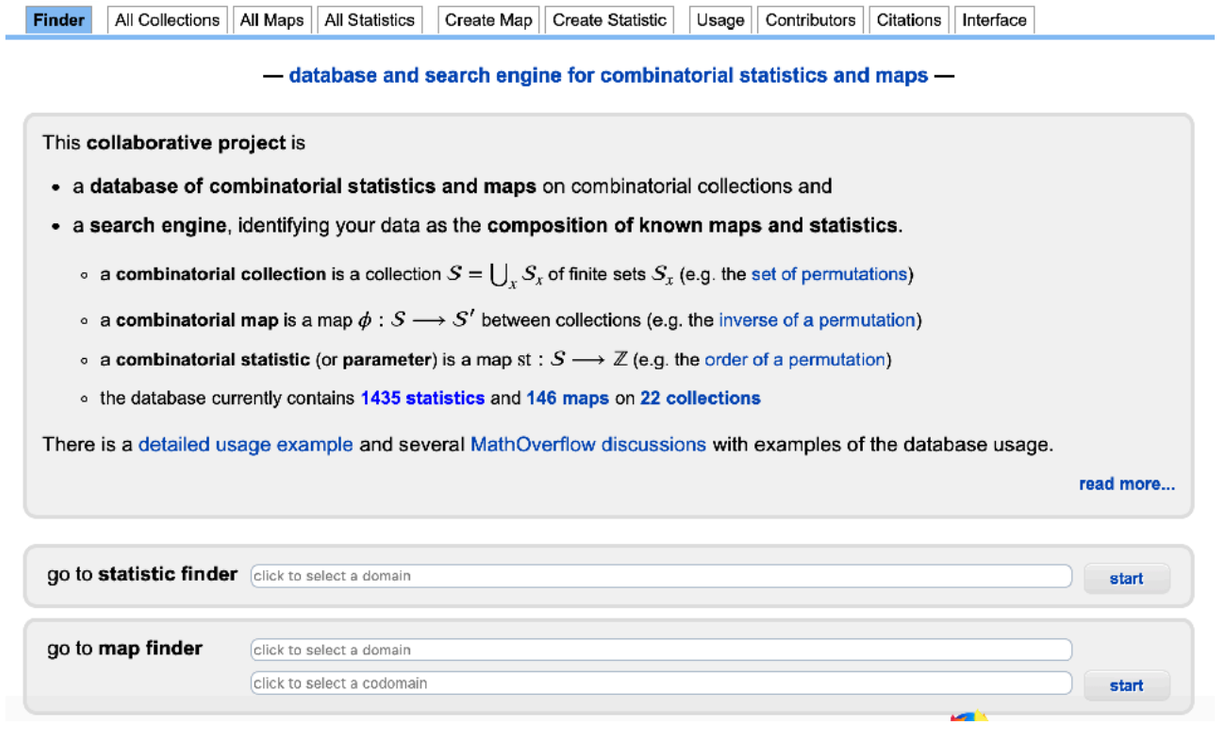


source: MathWorld

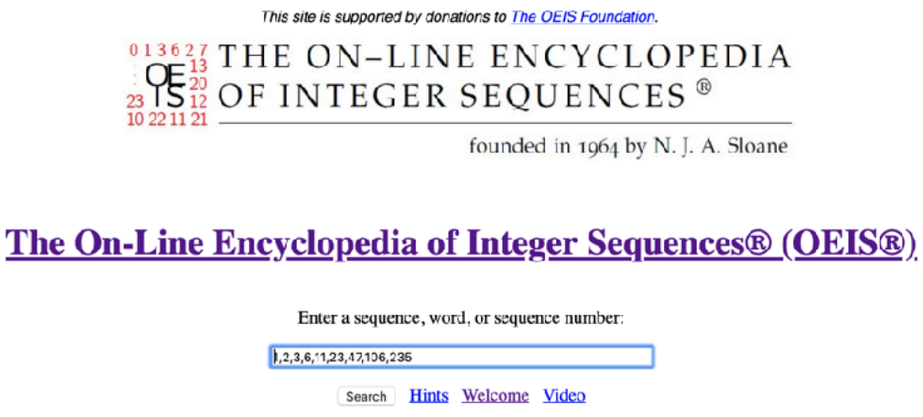
Some Famous Online Combinatorial Math Databases



lmfdb.org



findstat.org



oeis.org

The Other 80% (or more)

[N,i]	V	E	Tr	W?	B?	|AG|
C4[5, 1]	5	10	DT	W	NB	120
C4[6, 1]	6	12	DT	U	NB	48
C4[8, 1]	8	16	DT	U	Bip	(2^7)(3^2)
C4[9, 1]	9	18	DT	W	NB	72
C4[10, 1]	10	20	DT	U	NB	320
C4[10, 2]	10	20	DT	W	Bip	240
C4[12, 1]	12	24	DT	U	Bip	768
C4[12, 2]	12	24	DT	W	NB	48
C4[13, 1]	13	26	DT	W	NB	52
C4[14, 1]	14	28	DT	U	NB	(2^8)(7^1)
C4[14, 2]	14	28	DT	W	Bip	336
C4[15, 1]	15	30	DT	W	NB	60
C4[15, 2]	15	30	DT	W	NB	120
C4[16, 1]	16	32	DT	U	Bip	(2^12)
C4[16, 2]	16	32	DT	W	Bip	384
C4[17, 1]	17	34	DT	W	NB	68
C4[18, 1]	18	36	DT	U	NB	(2^10)(3^2)
C4[18, 2]	18	36	DT	W	Bip	144
C4[20, 1]	20	40	DT	U	Bip	(2^12)(5^1)
C4[20, 2]	20	40	DT	W	Bip	80
C4[20, 3]	20	40	DT	W	NB	320
C4[20, 4]	20	40	SS	U	Bip	(2^8)(3^1)(5^1)
C4[21, 1]	21	42	DT	W	NB	84
C4[21, 2]	21	42	DT	W	NB	336

Wilson, Potočnik; A Census of edge-transitive tetravalent graphs

- Graphs of order 4 to 300 (18 MB)
- Graphs of order 302 to 500 (66 MB)
- Graphs of order 502 to 600 (69 MB)
- Graphs of order 602 to 700 (84 MB)
- Graphs of order 702 to 800 (114 MB)
- Graphs of order 802 to 900 (147 MB)
- Graphs of order 902 to 1000 (183 MB)
- Graphs of order 1002 to 1050 (164 MB)
- Graphs of order 1052 to 1100 (113 MB)
- Graphs of order 1102 to 1150 (103 MB)
- Graphs of order 1152 to 1200 (234 MB)
- Graphs of order 1202 to 1250 (137 MB)
- Graphs of order 1252 to 1280 (131 MB)

Potočnik, Spiga, Verret; A census of small connected cubic vertex-transitive graphs

```
CubicVT:=[ [] : i in [1..1280]];

CubicVT[4,1] := Graph<4 | {{1,3}, {1,4}, {2,4}, {2,3}, {1,2}, {3,4}}>;

CubicVT[6,1] := Graph<6 | {{2,5}, {1,3}, {2,6}, {1,4}, {3,5}, {4,6}, {2,3},
{1,6}, {4,5}}>;

CubicVT[6,2] := Graph<6 | {{1,3}, {1,5}, {2,6}, {5,6}, {4,5}, {2,4}, {1,2},
{3,4}, {3,6}}>;

CubicVT[8,1] := Graph<8 | {{2,8}, {1,5}, {1,7}, {7,8}, {4,8}, {5,6}, {6,7},
{4,5}, {1,2}, {2,3}, {3,4}, {3,6}}>;

CubicVT[8,2] := Graph<8 | {{1,8}, {2,6}, {6,8}, {4,7}, {1,4}, {4,5}, {5,8},
{1,2}, {2,7}, {3,7}, {3,5}, {3,6}}>;

CubicVT[10,1] := Graph<10 | {{4,6}, {3,5}, {2,6}, {4,8}, {5,6}, {3,4}, {1,5},
{1,10}, {2,10}, {7,9}, {3,7}, {9,10}, {1,7}, {2,8}, {8,9}}>;

CubicVT[10,2] := Graph<10 | {{4,6}, {3,5}, {3,6}, {4,5}, {8,10}, {1,3}, {6,8},
{1,9}, {5,7}, {7,10}, {9,10}, {2,4}, {1,7}, {2,8}, {2,9}}>;

CubicVT[10,3] := Graph<10 | {{2,6}, {6,7}, {4,8}, {3,9}, {1,3}, {4,10}, {6,8},
{5,9}, {1,4}, {1,2}, {2,5}, {7,10}, {3,7}, {8,9}, {5,10}}>;

CubicVT[12,1] := Graph<12 | {{12,10}, {11,7}, {3,9}, {3,7}, {11,9}, {2,4},
{6,10}, {1,9}, {12,5}, {1,5}, {11,6}, {7,8}, {6,8}, {3,5}, {4,10}, {12,2},
{4,8}, {1,2}}>;
```

What mathematicians want

Authors

- dataset **hosting** with
- an easy way to submit a dataset

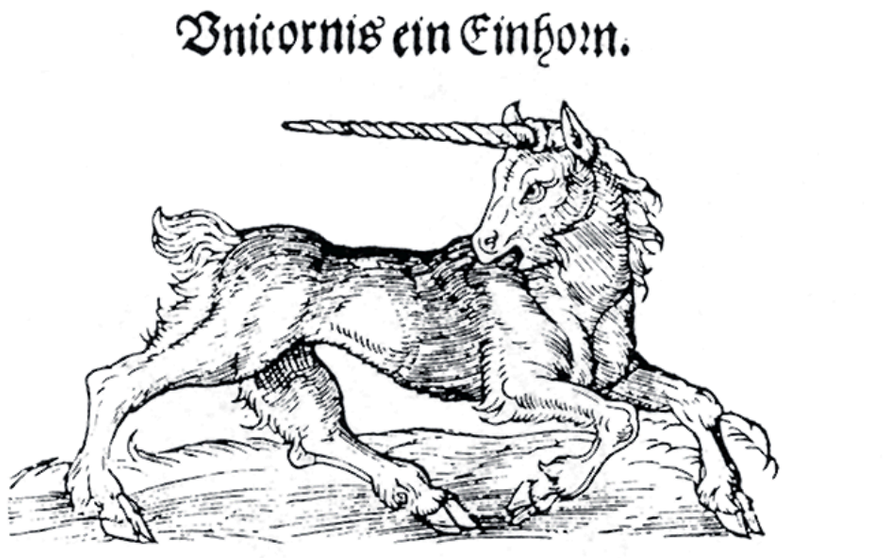
Users

- **look up**, **filter** subsets in a **math-level**, **searchable** interface (web, CAS)
- **run computations** (some kind of integration with CAS)

A solution?

standardize, reduce the number of
times we reinvent the wheel, and
give mathematicians what they want

hosting platform for math data



Albertus Magnus, *De animalibus*

Requirements

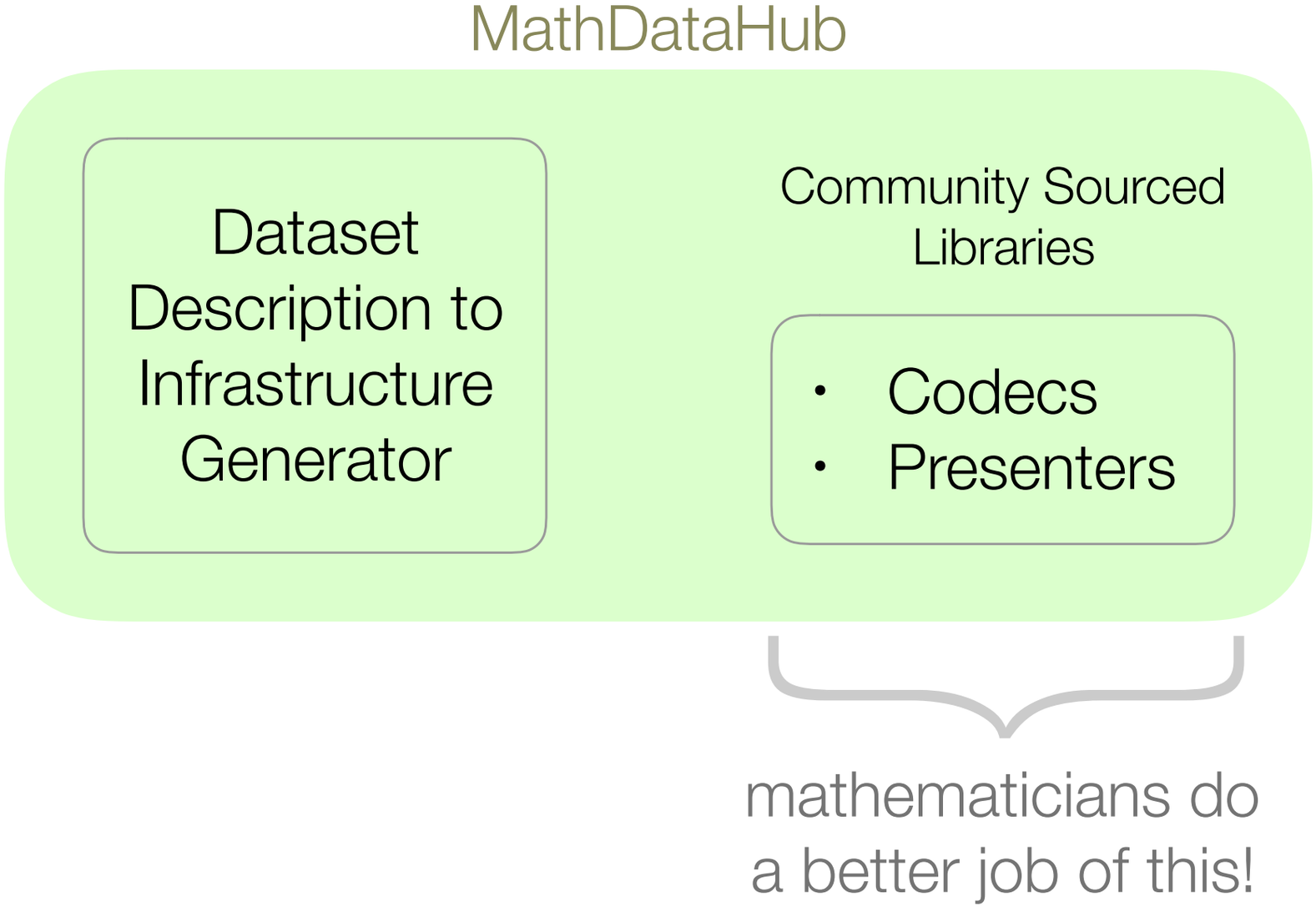
A **dataset description** language.

- *Generate* all the dataset specific infrastructure
- *Document* definitions of mathematical properties, encodings (database, CAS, ...)

Initially provide a **web interface**, build CAS APIs.

Enable system extension by **math-aware** features (mathematical presentation of data, mathematical query languages, data validation...)

An outline of a system



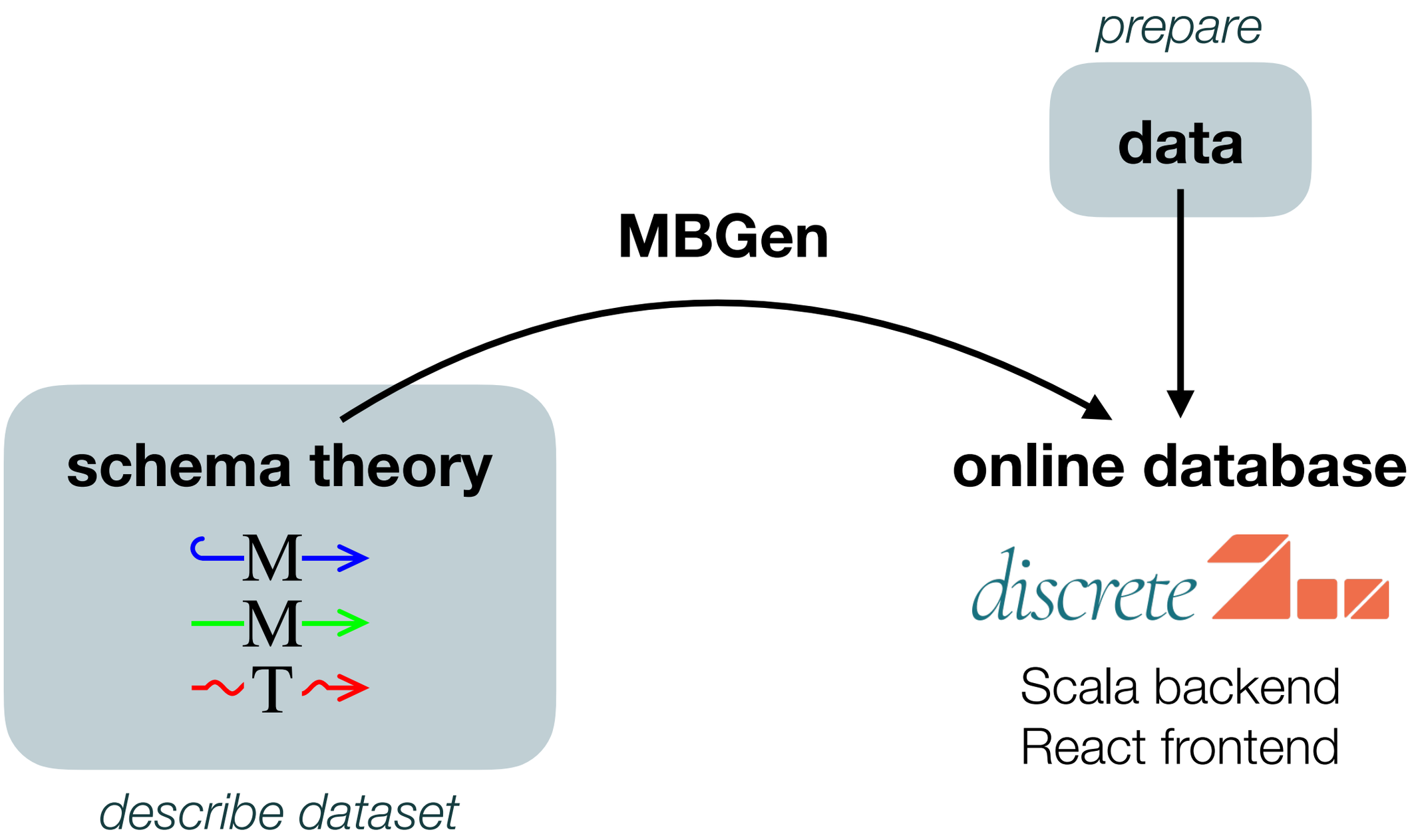
Conclusions + Future Work

- Provide the vital ingredient: a high-level data description framework MDDL. We already use this to produce a website stack.

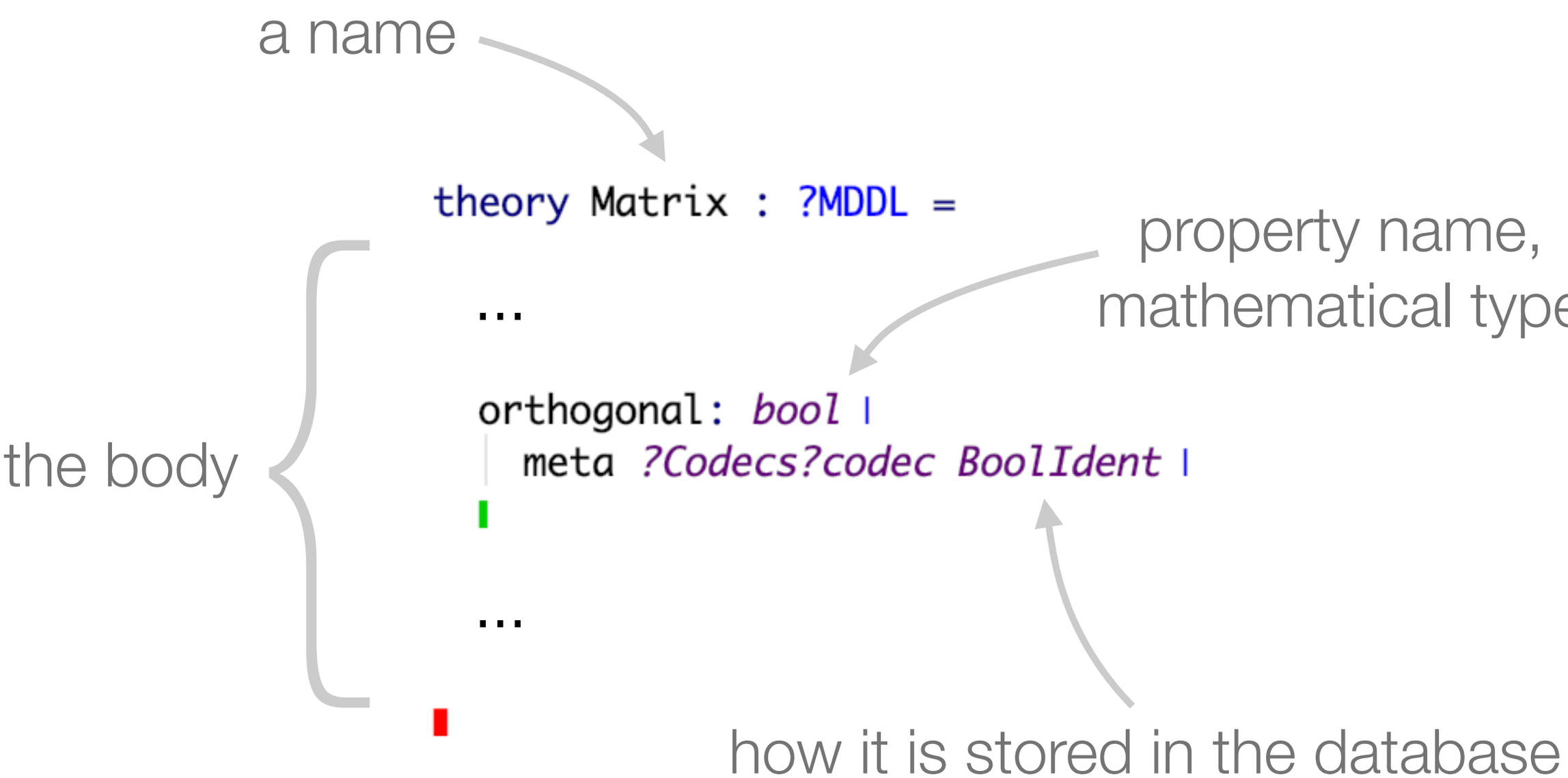
- Evaluate the system on existing datasets.

- Build a system that will enable math-level interfaces and features.

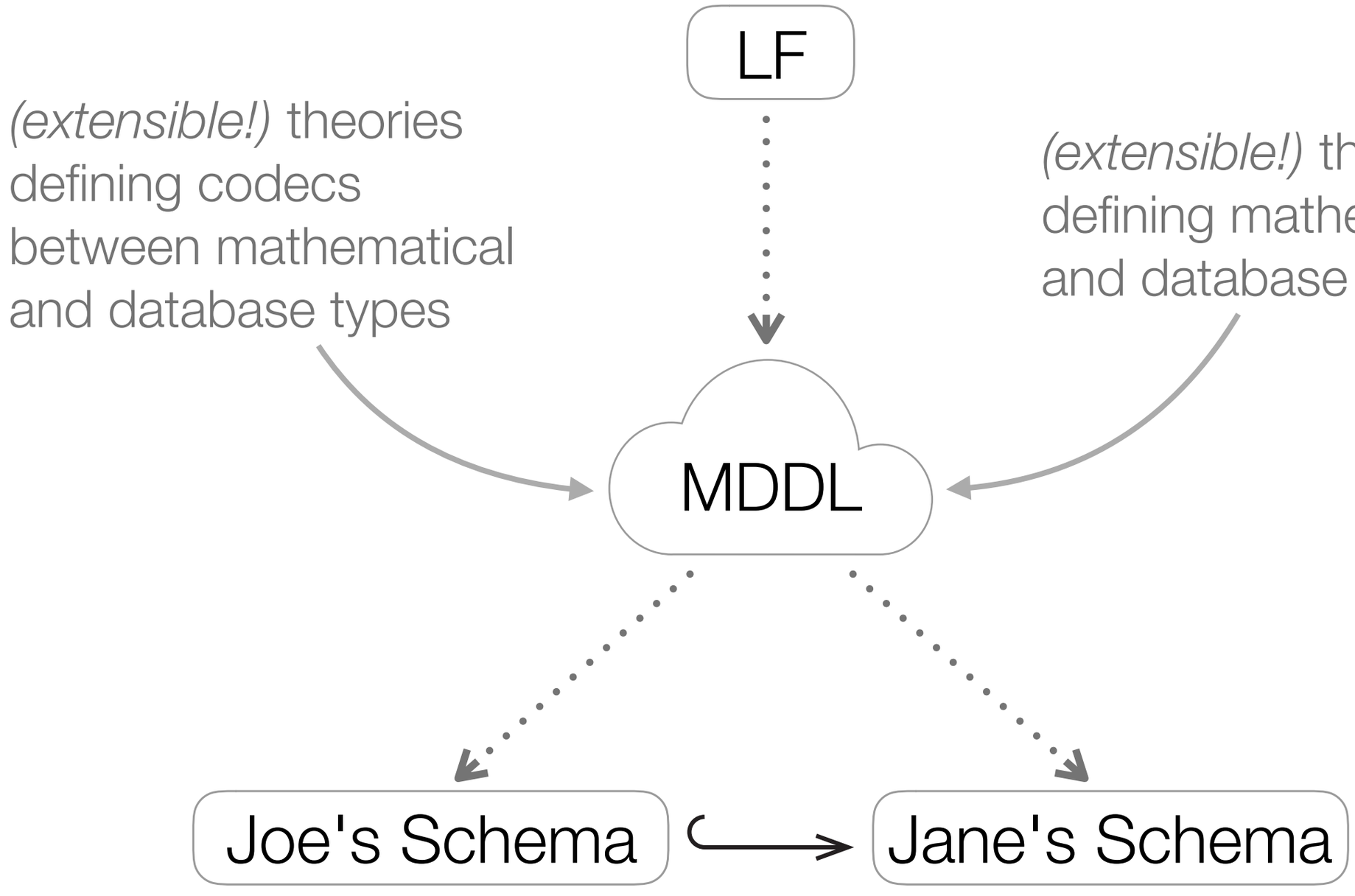
From a dataset to an online database



Anatomy of a schema theory



MDDL and Schema Theories



Search

Graphs | Maniplexes

- [ ] vertex transitive graphs
- [x] cubic vertex transitive graphs
- [ ] cubic arc transitive graphs

is partial cube

not is prism

chromatic index

clique number

connected components number

diameter

girth

Matches found: *4*

Display results

Choose columns

order	CVT	diameter	girth	is arc transitive	is cayley	is hamiltonian
20	7	5	6	true	false	true
24	11	6	4	false	true	true
48	29	9	4	false	true	true
120	60	15	4	false	true	true

```
namespace http://data.mathhub.info/schemas

theory MatrixS : ?MDDL =
  meta ?MDDL?schemaGroup "Joe"

  mat: matrix int 2 2
    meta ?Codecs?codec MatrixAsArray IntIdent
    tag ?MDDL?opaque
  

  trace: int
    meta ?Codecs?codec IntIdent
  

orthogonal: bool
  meta ?Codecs?codec BoolIdent
  

eigenvalues: list int
  meta ?Codecs?codec ListAsArray IntIdent
  tag ?MDDL?opaque
  
```