OpenDreamKit Work Package 6
Data/Knowledge/Software-Bases

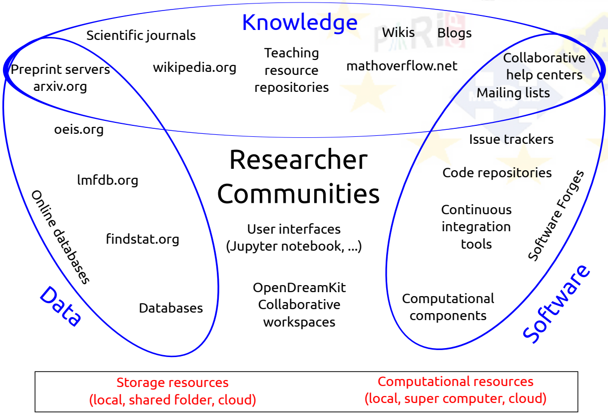
Michael Kohlhase
FAU Erlangen-Nürnberg
http://kwarc.info/kohlhase

OpenDreamKit Final Review, Luxembourg, October 30. 2019

1 Introduction

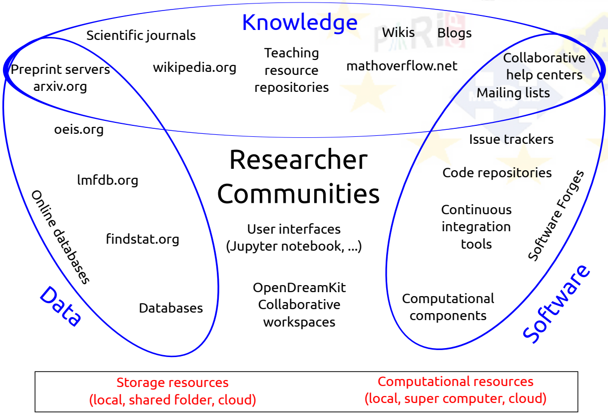
Background: WP6 (Data/Knowledge/Software-Bases)

- From the Proposal:



Background: WP6 (Data/Knowledge/Software-Bases)

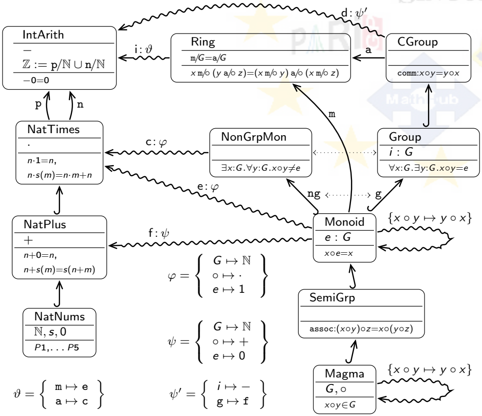
- From the Proposal:



- Proposed Focus: Supply this data to VRE components in an integrated fashion programmatically

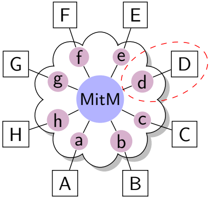
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation



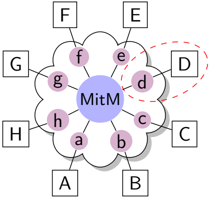
Results of the WP6 Workshops: Semantic Interoperability

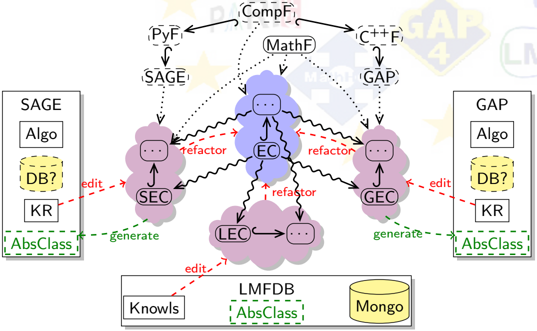
- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.



Results of the WP6 Workshops: Semantic Interoperability

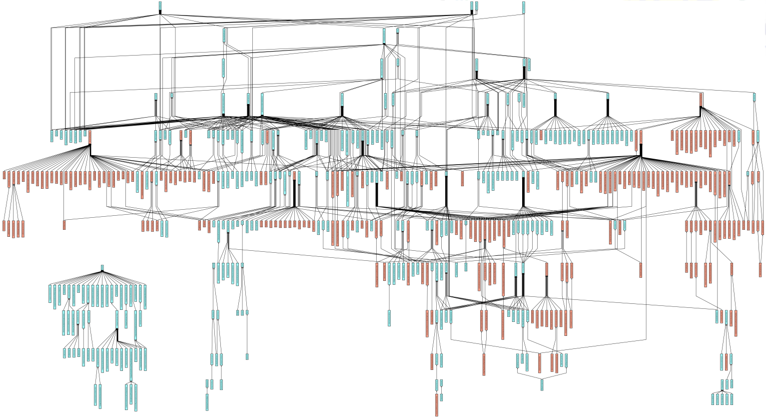
- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.





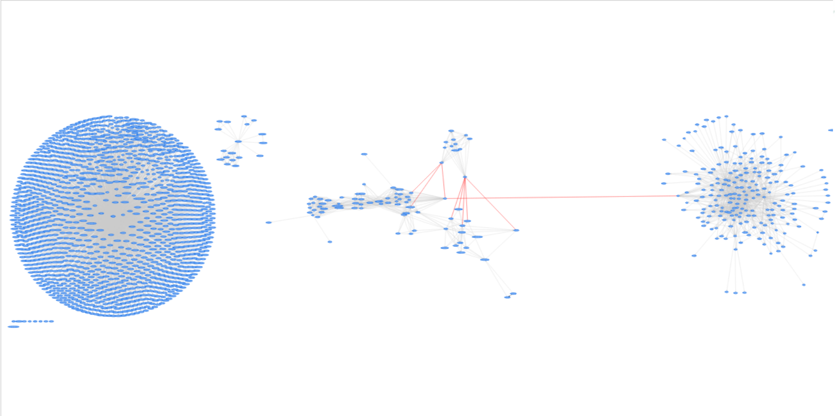
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)



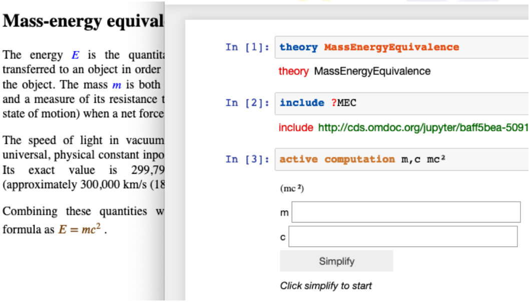
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology



Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug '17): Active, Structured Documents



Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug '17): Active, Structured Documents
  - WS in Cernay (May '18): Interfacing Systems

Sage

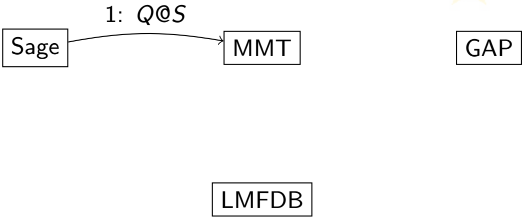
MMT

GAP

LMFDB

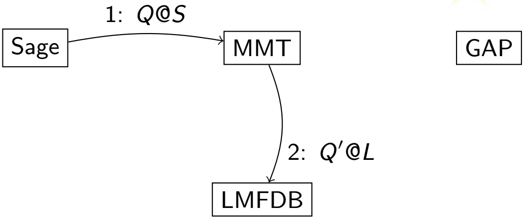
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug '17): Active, Structured Documents
  - WS in Cernay (May '18): Interfacing Systems



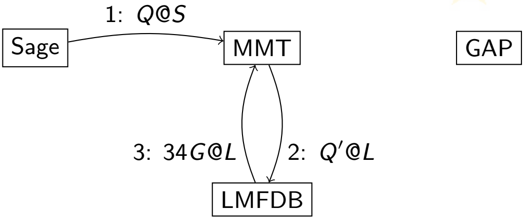
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug '17): Active, Structured Documents
  - WS in Cernay (May '18): Interfacing Systems



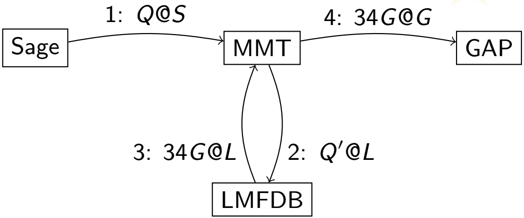
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep ’15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb ’16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June ’16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb ’17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug ’17): Active, Structured Documents
  - WS in Cernay (May ’18): Interfacing Systems



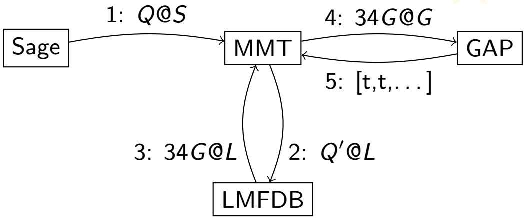
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug '17): Active, Structured Documents
  - WS in Cernay (May '18): Interfacing Systems



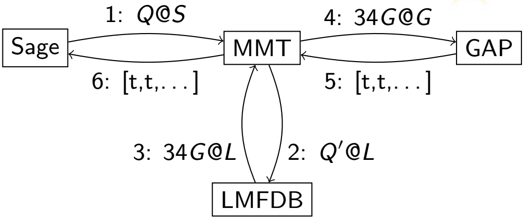
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug '17): Active, Structured Documents
  - WS in Cernay (May '18): Interfacing Systems



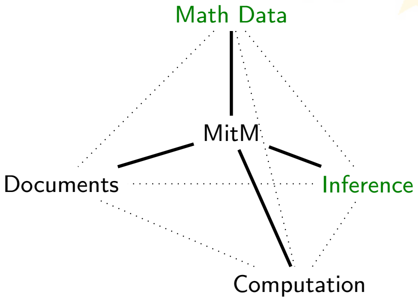
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug '17): Active, Structured Documents
  - WS in Cernay (May '18): Interfacing Systems



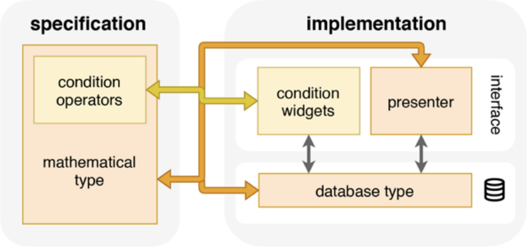
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug '17): Active, Structured Documents
  - WS in Cernay (May '18): Interfacing Systems
  - WS at FLoC Oxford (July '18): Aspects of Modular Knowledge



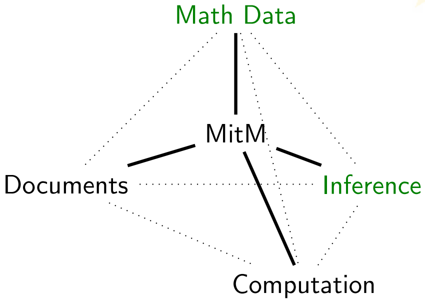
Results of the WP6 Workshops: Semantic Interoperability

- The WP6 group had a series of workshops
  - Kickoff in Paris (Sep '15): strategies for joint knowledge representation
  - WS in St. Andrews (Feb '16): Math in the Middle Arch. for System Interop.
  - WS in Bremen (June '16): GAP/SageMath API Content Dictionaries (CDs)
  - WS in Berlin (Feb '17): Math-in-the-Middle Ontology
  - WS in Oslo (Aug '17): Active, Structured Documents
  - WS in Cernay (May '18): Interfacing Systems
  - WS at FLoC Oxford (July '18): Aspects of Modular Knowledge
  - WS in Cernay (August '19): Math Data Workshop



WP6 Focus in the Final Review Period

- Inference $\rightsquigarrow$ the Isabelle Library and MitM
- Mathematical Data
  - (Semantic) Interoperability with Mathematical Data (Math-in-the-Middle)
  - Strengthening Organization via stronger Schemata (LMFDB)
  - Collecting mathematical Data during computation (Persistent Memoization)
- Data and Inference are a central part of “doing mathematics”.



2 Extending OpenDreamKit (MitM) to Inference

Integrating MitM with Theorem Proving – Isabelle Library

- New Task 6.11: *Isabelle Case Study* (last Amendment)
- Idea: Math uses a mixture of computation and and proving.
- Isabelle: One of the most mature and widely used proof assistants
  - 82 out of Wiedijk's top 100 math theorems formally proved
  - L4 microkernel verification: $> 10^5$ loc
  - Archive of Formal Proof
    > 300 authors, > 500 articles, $> 10^5$ lemmas, $> 10^6$ loc
- Subcontract: Collaboration with Makarius Wenzel (main Isabelle developer)
  Serialize Isabelle libraries in exchange formats (OMDoc/MMT) ($\approx 6 + 4$PM)
  - input
    - $> 10^4$ theories/locales
    - $> 10^6$ definitions and theorems
    - 135 MB uncompressed text files
  - output (without proofs)
    - 206 MB compressed OMDoc (37.5 GB uncompressed)
    - $> 10^8$ RDF triples
  - run time: 12 hours with 8 CPU cores, 50 GB memory

3 MathHub Data – your dataset, but FAIR

FAIR Research Data:The Next Big Thing

- **Definition 3.1.** Research data is recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings.
- Background: Virtually all scientific funding agencies now require some kind of research data strategy (tendency: getting stricter)

FAIR Research Data:The Next Big Thing

- **Definition 3.1.** Research data is recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings.
- Background: Virtually all scientific funding agencies now require some kind of research data strategy (tendency: getting stricter)
- **Definition 3.2 (Gold Standard Criteria).** Research data has to be FAIR, i.e.
  - Findable: easy to identify and find for both humans and computers, e.g. with metadata that facilitate searching for specific datasets,
  - Accessible: stored for long term so that they can easily be accessed and/or downloaded with well-defined access conditions, whether at the level of metadata, or at the level of the actual data,
  - Interoperable: ready to be combined with other datasets by humans or computers, without ambiguities in the meanings of terms and values,
  - Reusable: ready to be used for future research and to be further processed using computational methods.

FAIR Research Data: The Next Big Thing

- **Definition 3.1.** Research data is recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings.
- Background: Virtually all scientific funding agencies now require some kind of research data strategy (tendency: getting stricter)
- **Definition 3.2 (Gold Standard Criteria).** Research data has to be FAIR, i.e.
  - Findable: easy to identify and find for both humans and computers, e.g. with metadata that facilitate searching for specific datasets,
  - Accessible: stored for long term so that they can easily be accessed and/or downloaded with well-defined access conditions, whether at the level of metadata, or at the level of the actual data,
  - Interoperable: ready to be combined with other datasets by humans or computers, without ambiguities in the meanings of terms and values,
  - Reusable: ready to be used for future research and to be further processed using computational methods.

  Questions: What does this mean for mathematics, in particular
- - What is mathematical research data?
  - How can we make mathematical data FAIR?

The Current Reality in Mathematical Practice

- 80% of the datasets are not FAIR. (here are three silos)

[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
C4[17, 1]	17	34	DT	W	NB	68	4	1	0	4
C4[18, 1]	18	36	DT	U	NB	(2^10)(3^2)	512	128	1	4
C4[18, 2]	18	36	DT	W	Bip	144	8	2	2	4
C4[20, 1]	20	40	DT	U	Bip	(2^12)(5^1)	(2^10)	256	3	4
C4[20, 2]	20	40	DT	W	Bip	80	4	1	1	4
C4[20, 3]	20	40	DT	W	NB	320	16	4	1	4
C4[20, 4]	20	40	SS	U	Bip	(2^8)(3^1)(5^1)	384	96	0	4
C4[21, 1]	21	42	DT	W	NB	84	4	1	2	4
C4[21, 2]	21	42	DT	W	NB	336	16	4	2	3

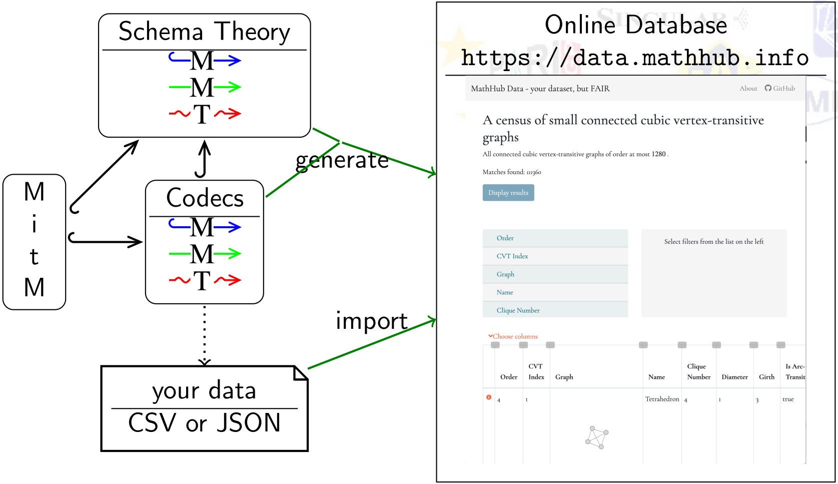
- Graphs of order 4 to 300 (18 MB)
- Graphs of order 302 to 500 (66 MB)
- Graphs of order 502 to 600 (69 MB)
- Graphs of order 602 to 700 (84 MB)
- Graphs of order 702 to 800 (114 MB)
- Graphs of order 802 to 900 (147 MB)
- Graphs of order 902 to 1000 (183 MB)
- Graphs of order 1002 to 1050 (164 MB)
- 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)

```
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}}>;
```

- Idea: Provide semantic hosting of all of these.

MathHub Data in a Nutshell

- MathHub Data:



- Community Resource: MitM and Codecs,
- Dataset: data in JSON, provenance, and schema theory.

Codecs: Encoding and Decoding Database Values

- **Definition 3.3 (Codec).** A codec consists of two functions that translate between semantic types and realized types.

Codecs

codec : $\mathtt{type} \to \mathtt{type}$	
StandardPos : $\mathtt{codec}\ \mathbb{Z}^+$ StandardNat :$\mathtt{codec}\ \mathbb{N}$ StandardInt :$\mathtt{codec}\ \mathbb{Z}$	JSON number if small enough, else JSON string of decimal expansion
IntAsArray :$\mathtt{codec}\ \mathbb{Z}$ IntAsString :$\mathtt{codec}\ \mathbb{Z}$	JSON List of Numbers JSON String of decimal expansion
StandardBool :$\mathtt{codec}\ \mathbb{B}$ BoolAsInt :$\mathtt{codec}\ \mathbb{B}$	JSON Booleans JSON Numbers 0 or 1
StandardString :$\mathtt{codec}\ \mathbb{S}$	JSON Strings

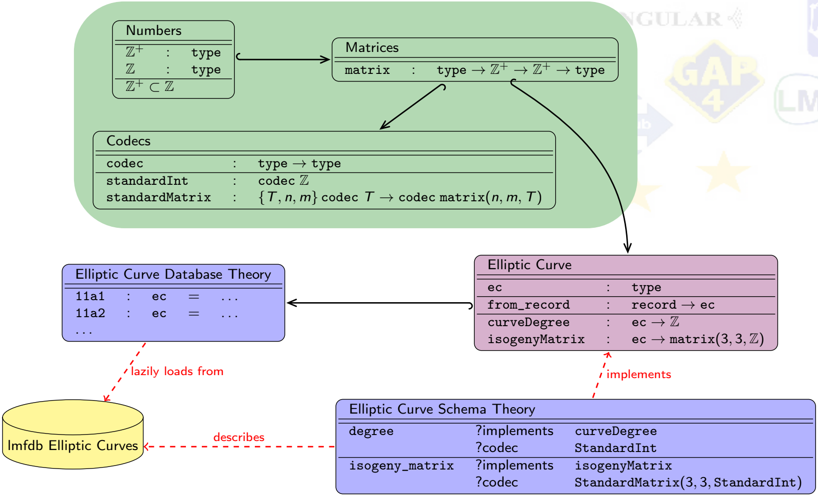
- StandardInt decodes 1 into the float `1`, but 2^{54} into the string `"18014398509481984"`

Elliptic Curve Code Operators


```
{
  "degree": 1,
  "x-coordinates_of_integral_points": "[5,16]",
  "isogeny_matrix": [[1,5,25],[5,1,5],[25,5,1]],
  "label": "11a1",
  "_id": "ObjectId('4f71d4304d47869291435e6e')",
  ...
}
```


- Matrix in the isogeny_matrix field
  - $$\begin{bmatrix} 1 & 5 & 25 \\ 5 & 1 & 5 \\ 25 & 5 & 1 \end{bmatrix}$$
  - represented as [[1,5,25],[5,1,5],[25,5,1]]

Our approach: Virtual Theories



MathHub Data (MHD) State of Play

- First working prototype since August 2019 (https://data.mathhub.info)
- Six datasets provided by the community (more in the pipeline)
  - Graphs, Maniplexes, Polyhedra, Additive Bases, small Groups...
  - together $\sim 13M$ Math Objects, 10 to 20 properties per object

  Mathematical variety sufficient to validate the system design.
- Wow: The DB researchers are very interested in the DB aspects (complex objects)
- Combinatorics community is very interested (Math Data WS $\leadsto$ 2020)
- Future: Scaling, Services, Community Building
  - Dataset submission process (metadata, descriptions, provenance, ...)
  - Working towards a "Journal of Mathematical Data" based on MHD
  - Semantic internal references via views.

Come to the MathHub Data Demo

MathHub Data - your dataset, but FAIR

About GitHub

A census of small connected cubic vertex-transitive graphs

All connected cubic vertex-transitive graphs of order at most 1280 .

This dataset has 111360 objects.

Matches found: *164*

More about this dataset

Display results

Available conditions

- Order
- CVT Index
- Graph
- Name
- Clique Number

Active conditions

- Order<50
- Clique Number>=2

Choose columns

	Order	CVT Index	Graph	Name	Clique Number	Diameter	Girth	Is Arc-Transitive	Is Bipartite	Is Cayley	Is Distance Regular	Is Distance Transitive	Is Edge-Transitive	Is Hamiltonian	Is Partia Cube
	4	1		Tetrahedron	4	1	3	true	false	true	true	true	true	true	false

4 Persistent Memoization

Persistent Memoization in Python and GAP

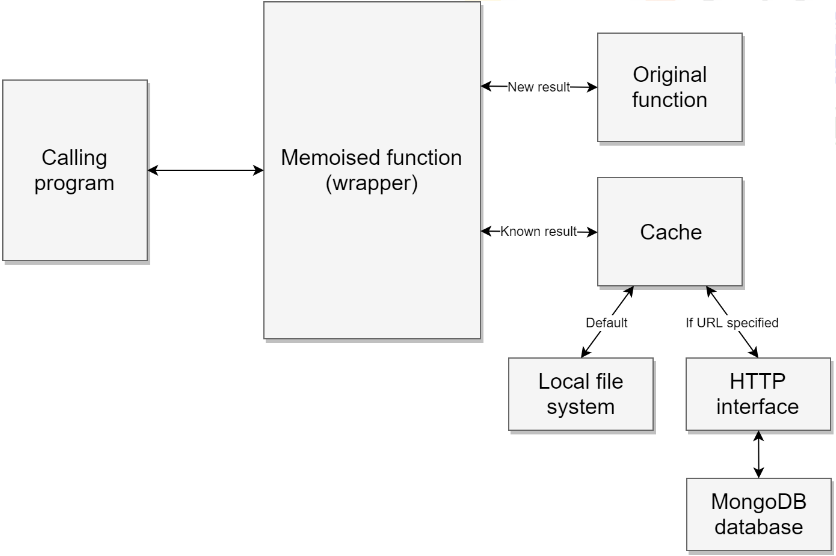
- What is memoisation?
  - Store program results in a permanent cache when they are computed
  - Retrieve these results from the cache later instead of recomputing
  - Cache can be local or online

Example 4.1 (Persistent Memoization in GAP/python).

```
In [8]:  double := x -> x * 2;;
         memo_double := MemoisedFunction(double);;

In [9]:  memo_double(4);

         #I  Memo key: [ 4 ]
         #I  Key unknown.  Computing result...

Out[9]:  8

In [10]: memo_double(4);

         #I  Memo key: [ 4 ]
         #I  Key known!  Loading result from cache...

Out[10]: 8
```

```
In [11]:
@persist(hash=lambda k: '%s to the %s' % (k[0][1], k[1][1]),
         pickle=str,
         unpickle=int)
def pow(x,y):
    return x**y

In [11]:
@persist(hash=lambda k: '%s to the %s' % (k[0][1], k[1][1]),
         pickle=str,
         unpickle=int)
def pow(x,y):
    return x**y
```

Persistent Memoization in Python and GAP



Persistent Memoization in Python and GAP

Persistent Memoization in Python and GAP

- Advantages
  - Avoids re-running programs that are guaranteed to return the same answer
  - Allows us to create an archive of results that can be used for other purposes
  - Share results betweeen users, locations, and even programming languages

5 Recommendations, Deliverables, KPIs, Lessons

Recommendations

- **Recommendation 7**. *To develop a comic explaining the MitM approach*.
  - The comic has been published on: `https://github.com/OpenDreamKit/OpenDreamKit.github.io/blob/master/public/images/use-cases/MitM.png`.
  - It has already been used in the MitM use case description at `https://opendreamkit.org/2018/05/16/lmfdb-usecase/`, in conference presentations and posters.
- **Recommendation 8**. *To disseminate the Adoption by Logipedia of the MitM principle of integrating (logical) systems by aligning concepts*.
  - We have made a blog post about this, see `https://opendreamkit.org/2019/01/24/logipedia/`

Deliverables in WP6

- All Deliverables were delivered (mostly on time)

Task	Name	RP1	RP2	RP3
T6.1	Search			
T6.2	Survey	D6.1		
T6.3	DKS-Design	D6.2 D6.3		
T6.4-8	Case Studies		D6.5 D6.8	
T6.9	Memoization			D6.9
T6.10	Math Search			D6.10
T6.11	Isabelle Lib			D6.11

KPIs and Deliverables for WP6

- The Math-in-the-Middle Ontology (largely unchanged from last time)
  - MitM-connected Systems: four (GAP, Sage, LMFDB, Singular) (See D6.5)
  - Formal MitM Ontology: 60 files, 3000 LoF, 500 commits (See D6.8)
  - Informal MitM Ontology: 900 theories, 1900 concepts in English, German, (Chinese, Romanian)
  - MitM System API Theories (GAP, Sage, LMFDB, Singular): 1.000+ Theories, 22.000 Concepts.
  - Isabelle Library: $> 10^5$ lemmas, $> 10^6$ loc
  - Heavy interest by the theorem proving community about MitM Ontology
  - Logipedia (`http://logipedia.science`) adopts the MitM principle of integrating (logical) systems by aligning concepts.

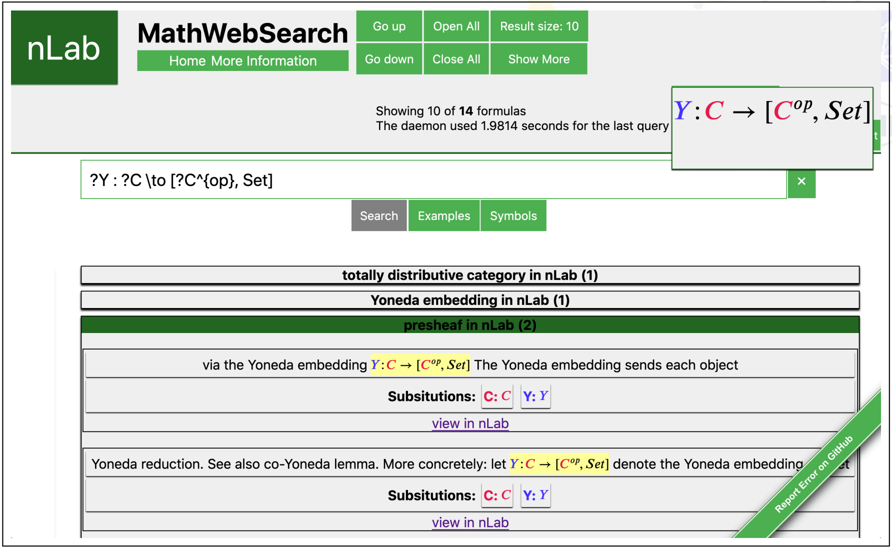
KPIs and Deliverables for WP6

- The Math-in-the-Middle Ontology (largely unchanged from last time)
  - MitM-connected Systems: four (GAP, Sage, LMFDB, Singular) (See D6.5)
  - Formal MitM Ontology: 60 files, 3000 LoF, 500 commits (See D6.8)
  - Informal MitM Ontology: 900 theories, 1900 concepts in English, German, (Chinese, Romanian)
  - MitM System API Theories (GAP, Sage, LMFDB, Singular): 1.000+ Theories, 22.000 Concepts.
  - Isabelle Library: $> 10^5$ lemmas, $> 10^6$ loc
  - Heavy interest by the theorem proving community about MitM Ontology
  - Logipedia (`http://logipedia.science`) adopts the MitM principle of integrating (logical) systems by aligning concepts.
- MathHub (Front-End re-implemented 2018/19)
  - 327 archives $\sim 61GB$; 25 archives in web UI; $\sim 2.5GB$

KPIs and Deliverables for WP6

- The Math-in-the-Middle Ontology (largely unchanged from last time)
  - MitM-connected Systems: four (GAP, Sage, LMFDB, Singular) (See D6.5)
  - Formal MitM Ontology: 60 files, 3000 LoF, 500 commits (See D6.8)
  - Informal MitM Ontology: 900 theories, 1900 concepts in English, German, (Chinese, Romanian)
  - MitM System API Theories (GAP, Sage, LMFDB, Singular): 1.000+ Theories, 22.000 Concepts.
  - Isabelle Library: $> 10^5$ lemmas, $> 10^6$ loc
  - Heavy interest by the theorem proving community about MitM Ontology
  - Logipedia (`http://logipedia.science`) adopts the MitM principle of integrating (logical) systems by aligning concepts.
- MathHub (Front-End re-implemented 2018/19)
  - 327 archives $\sim 61GB$; 25 archives in web UI; $\sim 2.5GB$
- MathHub Data (new, since August 2019)
  - $12M$ Math Objects with ~ 15 properties, $\sim 80GB$ in DB.
  - 4/6 data sets provided externally (four groups/researchers).

Come to the MathWebSearch Data (n-Category Lab)



Lessons Learnt: WP6 (Data/Knowledge/Software)-Bases

- Generally: OpenDreamKit was a tremendous opportunity to rethink Math Software Infrastructure
  - Freedom to think/conceptualize/prototype/evaluate/scale for 4 years
  - A common non-trivial infrastructure goal to evaluate (VRE toolkit)

Lessons Learnt: WP6 (Data/Knowledge/Software)-Bases

- Generally: OpenDreamKit was a tremendous opportunity to rethink Math Software Infrastructure
  - Freedom to think/conceptualize/prototype/evaluate/scale for 4 years
  - A common non-trivial infrastructure goal to evaluate (VRE toolkit)
- Problems Encountered: with taking the high road for system integration
  - MitM takes a large initial investment per system (Framework + mediator exist now)
  - mediator-based translation is relatively slow (but compilation possible)
  - correct/complete translations are possible by traditional programming (by trained mathematicians)
  - break-even point seems near 4 systems connected ($n^2 - n$ vs. $2n$ translations)

Lessons Learnt: WP6 (Data/Knowledge/Software)-Bases

- Generally: OpenDreamKit was a tremendous opportunity to rethink Math Software Infrastructure
  - Freedom to think/conceptualize/prototype/evaluate/scale for 4 years
  - A common non-trivial infrastructure goal to evaluate (VRE toolkit)
- Problems Encountered: with taking the high road for system integration
  - MitM takes a large initial investment per system (Framework + mediator exist now)
  - mediator-based translation is relatively slow (but compilation possible)
  - correct/complete translations are possible by traditional programming (by trained mathematicians)
  - break-even point seems near 4 systems connected ($n^2 - n$ vs. $2n$ translations)
- Fewer Problems encountered: for semantic mathematical data
  - semantic description of the dataset is a reasonable investment (Schema theories + JSON + Provenance)
  - BUT author gets a turnkey solution for their data sets! (first digitization)
  - AND the dataset is MitM-enabled. (both intra-MDH and with CAS)

OpenDreamKit Follow-Up Proposal: FAIRMat

- Call: European Research Infrastructures: Implementing the European Open Science Cloud (Deadline 29. 1. 2019)
- FAIRMat: FAIR Mathematical Data for the European Open Science Cloud
  - FAU Erlangen-Nürnberg (coordinator)
  - Univesité Paris Sud
  - Chalmers University of Technology
  - Univerza v Ljubljani
  - CAE Tech Limited
  - FIZ Karlsruhe – Leibniz Institute for Information Infrastructure
  - European Mathematical Society
- Work Areas:
  - WP2: Standardized data representation framework (deep FAIR)
  - WP3: Mathematical Services for the EOSC (e.g. search, programmatic APIs)
  - WP4: Data Sets for EOSC (Combinatorics, Algebra, Modelling)
  - WP5: Community Building
- Result: Cleared eligibility threshold well, not funded (too disciplinary)

Conclusion: What are we doing in WP6 in terms of a VRE

- SageMath and WP6 approach (Math-in-the-Middle; MitM) are both attempts at making a VRE Toolkit.
- SageMath is very successful, because integration is lightweight:
  - It makes no assumption on the meaning of math objects exchanged.
  - Restricts itself to master-slave integration of systems into SageMath.

  But there are safety, extensibility, and flexibility issues!

Conclusion: What are we doing in WP6 in terms of a VRE

- SageMath and WP6 approach (Math-in-the-Middle; MitM) are both attempts at making a VRE Toolkit.
- SageMath is very successful, because integration is lightweight:
  - It makes no assumption on the meaning of math objects exchanged.
  - Restricts itself to master-slave integration of systems into SageMath.

  But there are safety, extensibility, and flexibility issues!
- MitM tries to take the high road (make possible by OpenDreamKit)
  - Safety: by semantic (i.e. context-aware) objects passed.
  - Extensibility: any open-API system (i.e. with API CDs) can play.
  - Flexibility: full peer-to-peer possibilities. (future: service discovery)

  But we have to develop a whole new framework! (prototyped it)

Conclusion: What are we doing in WP6 in terms of a VRE

- SageMath and WP6 approach (Math-in-the-Middle; MitM) are both attempts at making a VRE Toolkit.
- SageMath is very successful, because integration is lightweight:
  - It makes no assumption on the meaning of math objects exchanged.
  - Restricts itself to master-slave integration of systems into SageMath.

  But there are safety, extensibility, and flexibility issues!
- MitM tries to take the high road (make possible by OpenDreamKit)
  - Safety: by semantic (i.e. context-aware) objects passed.
  - Extensibility: any open-API system (i.e. with API CDs) can play.
  - Flexibility: full peer-to-peer possibilities. (future: service discovery)

  But we have to develop a whole new framework! (prototyped it)
- Review Period3: Inference & Math Data
  - integrated Isabelle Library into MitM
  - Semanticizing LMFDB
  - Persistent Memoization
  - MathHub Data $\rightsquigarrow$ FAIR