

# COVER PAGE

**Title of Proposal:** MathVRE: Collaborative ecosystems for mathematical research and software development

**Date of preparation:** December 15, 2014

Participant no	Participant organisation name	Country
1 (Coordinator)	Université Paris Sud	FR
2	Logilab	FR
3	Université de Versailles Saint-Quentin	FR
4	Université Joseph Fourier	FR
5	Université Bordeaux	FR
6	University of Oxford	UK
7	Université of Sheffield	UK
8	Université of Southampton	UK
9	University of St Andrews	UK
10	University of Warwick	UK
11	University of Kaiserslautern	DE
12	University of Silesia	PL
13	Universität Zürich	CH
14	Simula Research Laboratory	NO

## Contents

<b>1</b>	<b>Excellence</b>	<b>2</b>
1.1	Objectives . . . . .	2
1.2	Relation to the Work Programme . . . . .	11
1.3	Concept and Approach . . . . .	13
1.4	Ambition . . . . .	16
<b>2</b>	<b>Impact</b>	<b>17</b>
2.1	Expected Impacts . . . . .	17
2.2	Measures to Maximise Impact . . . . .	18
<b>3</b>	<b>Implementation</b>	<b>20</b>
3.1	Work Plan — Work packages, deliverables and milestones . . . . .	20
3.2	Management Structure and Procedures . . . . .	48
3.3	Consortium as a Whole . . . . .	49
3.4	Resources to be Committed . . . . .	50
<b>4</b>	<b>Members of the Consortium</b>	<b>51</b>
4.1	Participants . . . . .	51
4.2	Third Parties Involved in the Project (including use of third party resources) . . . . .	66
<b>5</b>	<b>Ethics and Security</b>	<b>67</b>
5.1	Ethics . . . . .	67
5.2	Security . . . . .	67
	<b>References</b>	<b>68</b>

♦**EC Commentary:** Please follow the structure of this template when preparing your proposal. It has been designed to ensure that the important aspects of your planned work are presented in a way that will enable the experts to make an effective assessment against the evaluation criteria. Sections 1, 2 and 3 each correspond to an evaluation criterion for a full proposal.

*Please be aware that proposals will be evaluated as they were submitted, rather than on their potential if certain changes were to be made. This means that only proposals that successfully address all the required aspects will have a chance of being funded. There will be no possibility for significant changes to content, budget and consortium composition during grant preparation.*

*Page limit: The cover page, and sections 1, 2 and 3, together should not be longer than 70 pages. All tables in these sections must be included within this limit. The minimum font size allowed is 11 points. The page size is A4, and all margins (top, bottom, left, right) should be at least 15 mm (not including any footers or headers). If you attempt to upload a proposal longer than the specified limit, before the deadline you will receive an automatic warning, and will be advised to shorten and re-upload the proposal. After the deadline, any excess pages will be overprinted with a 'watermark', indicating to evaluators that these pages must be disregarded.*

*Please do not consider the page limit as a target! It is in your interest to keep your text as concise as possible, since experts rarely view unnecessarily long proposals in a positive light.♠*

## Outline of Project (for Proposers)

♠**TO DO:** *This is the place for various READMEs not included in the final submission*♠

### Mission statement for the grant

Our mission is to promote the next generation of community-developed open source software, databases, and services adapted to the needs of collaborative research in pure mathematics and applications.

Our research will cover a wide variety of aspects, ranging from software development models, user interfaces ♠**TO DO:** *virtual environments?*♠, deployment frameworks and novel collaborative tools, component architecture, design, and standardization of software ♠**TO DO:** *system?*♠ and databases, to links to publication, data archival and reproducibility of experiments, development models and tools, and social aspects.

It will consolidate Europe's leading position in computational mathematics and build on the remarkable success of the ecosystem of projects GAP, Python/Sage, Pari, Singular, LMFDB.

♠**TO DO:** *What do we mean by "new generation"*♠.

# 1 Excellence

**MathVRE** focuses on delivering easy to setup Virtual Research Environments, customizable to meet the diverse needs of collaborative research in pure mathematics and applications, and built out of a sustainable ecosystem of community-developed open software, databases, and services.

The remarkable emergence and success in the last decade(s) of a large ecosystem of community-developed computational software like e.g. GAP, Linbox, Pari, Sage, or Singular and of interactive scientific computing environments like IPython showcase the viability of community development models for such components.

The recent successes of the Virtual Research and Teaching Environment SageMathCloud (hosting more than 10k users and 100k projects after just one year) and of the online number theory database LMFDB showcase the strong need for **integrated solutions** enabling large-scale collaboration on **software**, **knowledge**, and **data**.

Yet setting up infrastructures like those mentioned above currently requires massive ad-hoc efforts. Challenges include portability, compatibility, performance, usability, reproducibility, not to mention the many social aspects involved in communities and ecosystems thereof.

♠**TO DO:** *Better explain the specific needs in maths, and why we focus on VRE bricks rather than VRE themselves, in the DIY spirit.*♠

A specific challenge in mathematics comes from the vast yet tightly connected array of concepts involved. Different groups of researchers may have radically different needs, workflows, and resources, calling for a highly modular and customizable VRE infrastructure.

**MathVRE** will attack all those challenges upfront, while consolidating Europe's leading position in this field.

## 1.1 Objectives

♠**EC Commentary:** *1-2 pages*♠ ♠**EC Commentary:** Describe the specific objectives for the project, which should be clear, measurable, realistic and achievable within the duration of the project. Objectives should be consistent with the expected exploitation and impact of the project (see section 2).♠

The specific aims of **MathVRE** are:

**Aim 1:** Improve the productivity of researchers by promoting collaborations based on mathematical **software**, **data**, and **knowledge**.

**Aim 2:** Make it easy for teams of researchers of any size to setup custom collaborative Virtual Research Environments tailored to their needs and workflows, supporting the entire life-cycle of computational work in mathematical research, from initial exploration to publication, teaching and outreach.

**Aim 3:** Identify and promote best practices in computational mathematical research, including making results easily reproducible, making software sustainable, reusable and easily accessible and sharing data in a semantically sound way. ♠**TO DO:** *insert the ecosystem keyword*♠

Our research will cover a wide variety of aspects, ranging from software development models, user interfaces ♠**TO DO:** *virtual environments?*♠, deployment frameworks and novel collaborative tools, component architecture, design, and standardization of software ♠**TO DO:** *system?*♠ and databases, to links to publication, data archival and reproducibility of experiments, development models and tools, and social aspects.

The concrete objectives of **MathVRE** are:

**Objective 1:** To develop and standardise an architecture allowing a range of mathematical and data and software components to be combined with off-the-shelf non-mathematical infrastructure to produce specialised VREs for different research communities. The architecture will take the form of standards documents and APIs equipped, where appropriate, with formal or informal mathematical semantics to ensure interactions are mathematically sound. This primarily addresses aim 2, thereby contributing to aims 1 and 3. ♠**TO DO:** *This is a bit long*♠

**Objective 2:** To develop open source core components for such VREs where existing software is not available. These components should support VREs running on a variety of platforms, including standard e-infrastructures. This primarily addresses Aim 2, thereby contributing to Aim 1 and 3.

♠**TO DO:** *Urgent: talk to EPCC re e-infrastructure standards*♠

♠**TO DO:** *SL Write detailed description*♠

**Objective 3:** Community building

Bring together the communities (IPython, Sage, Singular)

♠**TO DO:** *NT expand and write detailed description*♠

**Objective 4:** Update a range of existing open source mathematical software systems for seamless deployment and efficient execution within the VRE architecture of objective 1 This fulfills part of Aim 2.

♠**TO DO:** *SL: write detailed description*♠

**Objective 5:** Foster a sustainable ecosystem of interoperable open source components developed by overlapping communities, in particular by identifying and promoting software development best practices, and outsourcing development to larger communities whenever suitable. This fulfills part of Aim 3.

**Objective 6:** Explore the social aspects: how do researchers collaborate in Mathematics? What can be the role of Virtual Research Environments? This addresses part of Aim 3 and 2. ♠**TO DO:** *UM: write detailed description*♠ ♠**TO DO:** *SL: semantics from mathematical text?*♠

**Objective 7:** Identify and extend ontologies and standards to allow easy, safe and efficient storage, reuse, interoperation and sharing of rich mathematical data taking account of provenance and citability. This fulfills parts of Aim 2 and 3. ♠**TO DO:** *POD – write detailed description*♠

**Objective 8:** Demonstrate the effectiveness of Virtual Research Environment built on top of **MathVRE** components for a number of real-world use cases taken from different domains, or crossing previously little connected domains. This addresses part of Aim 2.

**Objective 9:** Effective Dissemination ♠**TO DO:** *VP – write detailed description*♠

## Detailed Descriptions of Objectives

### Objective 1: Virtual Research Environment Kit

Computational techniques have become a core asset for research in pure mathematics and its applications in the last decades. Mathematics communities have come together to develop powerful computational tools, such as GAP, Pari, SAGE or Singular, and valuable on-line services such as the Encyclopedia of Integer Sequences and the ATLAS of Group Representations. ♠**TO DO:** *cite*♠ In building these systems, mathematicians have gained strong experience in collaborative software development, with pioneering work and continuing leadership of Europe.

A number of approaches to linking these resources have been developed, such as the SCSCP protocol from the Framework 6 SCIENCE project, and the incorporation of a variety of free software tools in the SAGE system, but the overall model is still that of a single mathematician running programmes or interacting with a “notebook” page. The software provides little or no support for other aspects of mathematical research: collaboration, archival, reproducibility or linkage between programmes, data and publication. Databases are updated mainly by mathematicians directly, retaining no record of the source of new entries, and providing no way of referring to the actual version of the data used in a particular computation.

The first objective of this project is to design an architecture which will allow existing mathematical software systems (suitably updated), off-the shelf non-mathematical tools and a small number of new components to be flexibly combined to produce a VRE that will effectively support collaborative mathematical research throughout its entire life-cycle. This will include software APIs and standards, frameworks for assuring the semantic consistency of similar mathematical objects in different systems. It will be informed by the outputs of objective 6, ensuring that the VREs fit the ways that mathematicians actually work.

### Objective ??: ♠**TO DO:** *title*♠

Most of the direct mathematical capabilities of our software will come from existing open source mathematical systems. For instance we will use the power of the GAP Library for computational group theory or Pari for number

theory. Generic services such as storage, version control, authentication and resource accounting will come from off-the-shelf components building on standard infrastructures.

Nevertheless some new tools will be needed ♠**TO DO: what? or at least examples**♠

♠**TO DO: Keyword: flexible/versatile virtual environment**♠

Our research will cover a wide variety of aspects, ranging from software development models, user interfaces ♠**TO DO: virtual environments?**♠ deployment frameworks and novel collaborative tools, component architecture, design, and standardization of software components and databases, to links to publication, data archival and reproducibility of experiments, development models and tools, and social aspects. It will build on the remarkable success of the open source ecosystem and consolidate Europe's leading position in computational mathematics.

Following the call specifications, all software, data, and publications resulting from this proposal will be open.

### **Objective 5: A sustainable ecosystem of software components**

The success of large specialized software like Pari, Singular or GAP in the last decades has shown the viability of the academic open source development model for such. For a long time, it was bitterly debated whether this model would have any chance to scale to general purpose systems for pure mathematics. The rapid takeoff of Sage in the last 10 years has proven the viability of the “developed by users for users” model: despite its large community of about 150 active developers, it's running on a tiny specific budget, with most activities being funded indirectly by research grants that require specific development.

This was made possibly by reusing existing components whenever possible (e.g. hundreds of specialized open source math libraries, or the Python programming language with its developers tools and huge library), and spinning off software development (e.g. the Cython compiler) to larger communities whenever possible.

Yet, long term critical non mathematical features like portability, modularization, packaging, user interfaces, large data, parallelism, or outreach toward related software, have been lagging behind. Indeed they can hardly be implemented as a side product of research projects, and **need to be assigned to a few full time developers**. Regular funding is also needed to better structure the computational mathematics community in Europe and support its upcoming major widening through training, development workshops, exchanges, ...

**One purpose of this grant is to initiate this process and invest on the long run to reduce the recurrent needs.**

The principle is that, with the growth of the user base, a tiny number of institutions or companies will hire a full-time developer to support critical needs of their in-house research or development. Opportunities for such hiring are for example actively investigated at the Laboratoire de Recherche en Informatique. It should be noted that, at the scale of a large university, the cost of licenses for analogous commercial software can reach the same order of magnitude as that of a developer.

To reduce the number of required full time developers **MathVRE** will invest toward, factoring out joint needs, and outsourcing or spinning off more components to larger communities. For example, **MathVRE** will save much recurrent efforts to the mathematics community by providing a temporary boost to outsource the development of the user interface of each computational component to IPython and make IPython stand to the stringent needs of the community. Later on, thanks to its large user base both in academia and industry, IPython will continue to thrive without specific funding or major contributions from the mathematics community.

**MathVRE** will also foster the productivity within the ecosystem by investigating better collaboration processes between components, and identifying, sharing, and promoting software development best practices.

### **Objective 9: ♠**TO DO: Title**♠**

The success of any research software or service is strongly related to its ability to attract and convince a great amount of users. Our different communities (Sage, Gap, Pari, Singular...) have developed a solid experience and network. As an example, Sage has gathered thousands of users in less than 10 years. This was achieved thanks to a very strong community building philosophy, especially through the organization of Sage-Days all over the world. The first Sage-days happened in 2006, today we count 63 of them, including 10 for 2014 only, and also Sage Education days, Sage Bug days, Sage Doc days and more. Most of us, **MathVRE** project members, have been involved in these events either as organizers or participants and it appears as the most efficient way to promote our software. More precisely, our objective is to create constant dialogue with the different communities: frequent workshops, conferences, user groups, mailing lists. By building on existing tools, we intend to involve the communities in the development process itself in the spirit of open-source software.

We also intend to reach a larger crowd of researchers by cutting down non-research technical issues to access existing tools: building better documentation and tutorials, developing easy-to-install distributions, easy web and cloud access, better user interfaces, better interactions between different software. Doing so, our objective is to help the communities to grow themselves and interact together using our work.

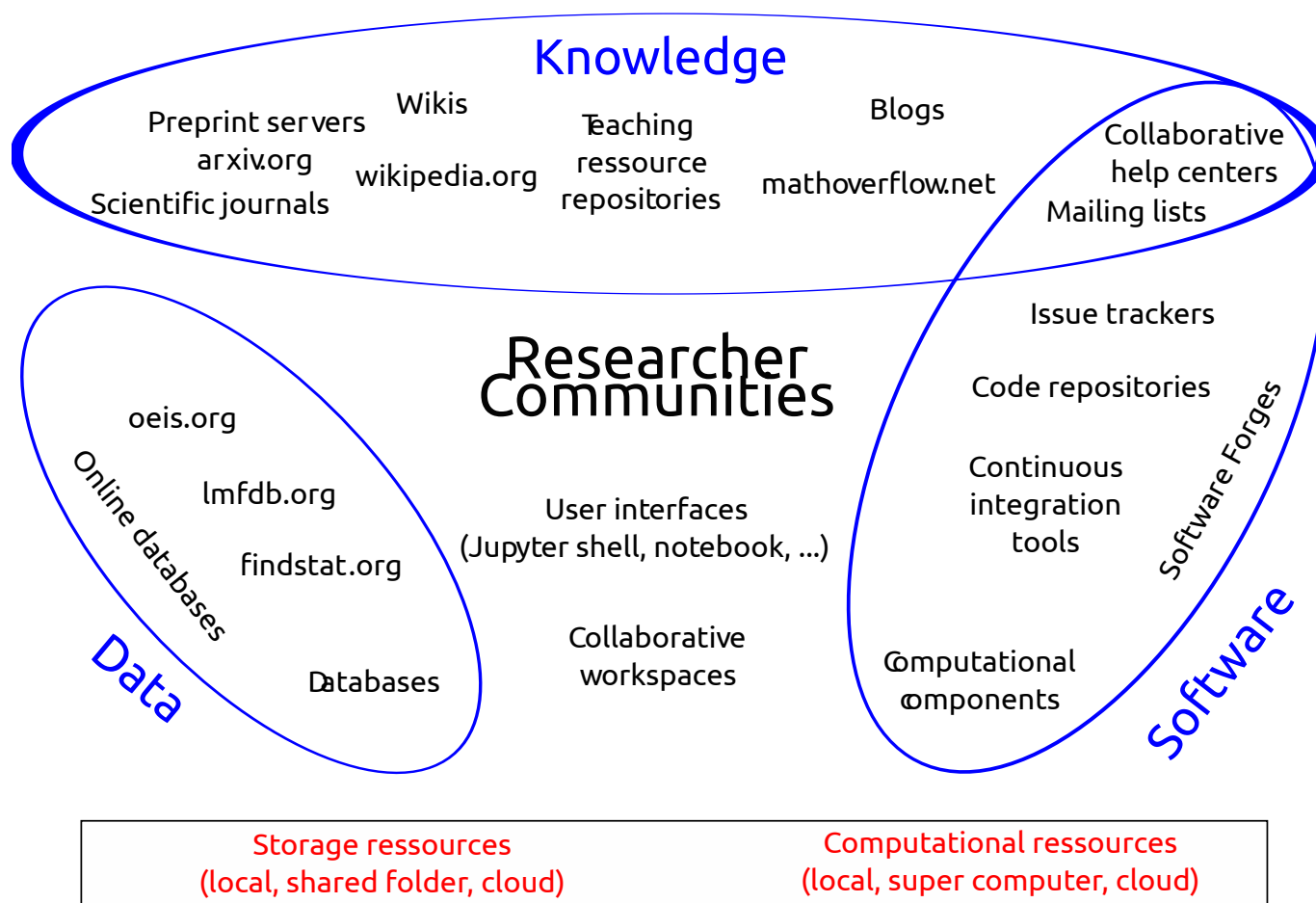


Figure 1: Virtual Research Environments for research in pure mathematics and applications.

♠**TO DO:** *the findstat link does not work for me, kerning looks extremely weird – POD♠*

♠**TO DO:** *lmfdb, oeis, and findstat have a strong knowledge component as well, with knowls and wikis, references, ...♠*

♠**TO DO:** *arxiv is not far from a database of knowledge♠*

♠**TO DO:** *NT: the purpose of Figure ?? is to give a quick sense of what Virtual Research Environments can be in our context, and a “big picture” for the project. A graphic artist friend of mine is going to help me improve it. I have collected here some material for her.*

### ***What we would like the “big picture” in Figure 1 to highlight:***

**This is a human centered project:** *At the core: researchers and communities thereof.*

**The three types of information:** *Software, Knowledge, Data (currently in blue)*

*How they interact:*

- *Knowledge help structure data and software (e.g. through ontologies)*
- *Software produce data*
- *Data is used by researchers to build knowledge*

**Physical resources:** *(currently in red)*

### **Virtual Research Environments**

- *Researchers in Math have a long tradition of collaborating on Software, Knowledge, and, up to some point, Data*



- *For this they use a variety of collaborative tools which form a loosely knit Virtual Research Environment.*
- **Aim 2:** *make it easy for subcommunities of researchers to setup custom collaborative work spaces / Virtual Research Environments tailored to their needs, by combining:*
  - *Computational resources*
  - *Storage resources*
  - *Computational software components*
  - *Databases*
  - *User interfaces*
  - *Wikis-Knowledge bases (true for findstat, LMFDB): quicker cycle for consolidation of information spread over papers/brains*

*Such VRE shall help them:*

- *collaboratively develop software (e.g. specialized libraries), data and knowledge (e.g. articles) for their research projects.*
- *contribute back this information to the larger community whenever relevant.*

## **Processes:**

*It would be interesting to depict the following processes. They are indeed about collaboration and sharing (and quality control), that is what **Aim 1** is to promote.*

### **Software development**

- *bug reports and enhancement requests emerge from the community, typically through collaborative help centers, and are posted on issue trackers.*
- *Design discussions occur on mailing lists and issue trackers.*
- *Researchers submit code to the code repositories.*
- *Quality control: the code is reviewed and tested by continuous integration tools.*
- *Finally the code integrated within computational components, and used by the community.*

*Researchers (as well as other users: teachers, engineers, ...) interact at each step of the process.*

### **Scientific publication**

- *researchers submit articles to journals and post them on preprint servers;*
- *the articles get reviewed by other researchers;*
- *finally they are distributed back to the community*

## ***A collection of links that might give some idea of the look and feel of our universe:***

### **Examples of (computational) components:**

- *IPython: <http://ipython.org/>*
- *GAP: <http://www.gap-system.org/>*
- *Singular: <http://www.singular.uni-kl.de/>*
- *Sage: <http://sagemath.org/>*
- *Pari/GP: <http://pari.math.u-bordeaux.fr/>*
- *Linbox: <http://www.linalg.org/>*

### **Examples of online collaborative tools**

- *Issue tracker: <http://trac.sagemath.org/timeline/>*
- *Code repository: <https://github.com/>*
- *Collaborative help center: <http://ask.sagemath.org/>*
- *Collaborative math site: <http://mathoverflow.net/>*

### Examples of online databases

- *Online databases:* <http://oeis.org/?language=french>
- *LMFDB:* <http://www.lmfdb.org/EllipticCurve/Q/14.a3>
- *Findstat:* <http://www.findstat.org/>

### Example of graphical material

- <http://boxen.math.washington.edu/home/nthiery/main2014.pdf>



### 1.1.1 Why collaborative development of open source software?

♠**TO DO:** *Language to be moved in some form into the Concept section*♠

From their early days, computers have been used in pure mathematics, either to prove theorems or, like the telescope for astronomers, to explore new theories. Major achievements include the proof of the four color theorem or ♠**TO DO:** *Nice flashy example?*♠. Usage has grown to the point that certain areas of mathematics now completely depend on experimental methods, with major efforts spent on software development. As the sophistication of the required computations increased, supported by the boom of the available computational power, it became vital to share those efforts at the scale of large research communities. European mathematicians have been pioneers and have grown a steady tradition of collaborative open source software development, with systems like GAP, Singular, or Pari/GP playing a major role for decades.

### 1.1.2 Importance of experimental tools in maths

♠**TO DO:** *Language to be moved in some form into the Concept section*♠

The field of computer algebra allows us to compute in and with a multitude of mathematical structures. It is interdisciplinary in nature, with links to quite a number of areas in mathematics, with applications in mathematics and other branches of science and engineering, and with constantly new and often surprising developments. Quite a number of these developments, in fact the creation of whole subareas of the field, have been initiated by European researchers who made crucial contributions at all levels. These include the design of fundamental algorithms, the development of major computer algebra systems, applications of the computational methods in various fields, and the creation of widely used databases.

Particular fruitful interactions unfold between computer algebra and algebraic geometry, number theory, combinatorics and group theory. Algebraic algorithms open up new ways of accessing subareas of these key disciplines of mathematics, and they are fundamental to practical applications of the disciplines. Conversely, challenges arising in algebraic geometry, number theory, combinatorics and group theory quite often lead to algorithmic breakthroughs which, in turn, open the door for new theoretical and practical applications of computer algebra.

Based on exact computer aided calculations, the experimental method has now been added to the toolbox of the pure mathematician. Experiments lead to new conjectures which may have a deep impact on the future development of mathematics. An outstanding example is the Birch and Swinnerton-Dyer conjecture which is one of the Clay Millennium Problems. Databases relying on computer calculations such as the Small Groups Library or the Modular Atlas in group and representation theory provide indispensable tools for researchers. A constructive way of understanding proofs of deep theorems yields algorithmic tools to deal with highly abstract concepts. These tools make the concepts available to a broader class of researchers, with many potential applications. A prominent example from algebraic geometry is the desingularization theorem of Hironaka, for which Hironaka won the Fields Medal, and its algorithmization by Villamayor.

Spectacular theoretical breakthroughs such as the recent complete resolution of Serre's conjectures, directly inspired by Wiles' proof of Fermat's last theorem, are based on interdisciplinary approaches. Current developments on the algorithmic side allow one to conquer crossconnections between different areas of mathematics also computationally and, thus, to arrive at cutting-edge applications which previously were inconceivable.

Other chunks of language:

- Mathematicians have a strong tradition of sharing knowledge openly (arxiv, Wikipedia, ...).
- Mathematicians have been building and sharing databases for a long while; the needs for such is growing tremendously, and the process needs to be streamlined.
- Specific situation of maths w.r.t. Data:
  - More often than not data is the result of a computation (and not e.g. an experiment). The role of databases is thus primarily to store results for later reuse (persistent caching), and enable searches. Because of this, many issues (semantic, ontologies, reproducibility, ...) are to be treated upstream at the level of software rather than data.
  - extreme reification in mathematics makes classical ontologies techniques/RDF impractical
  - interlinking very high

- several alternate and defining description of same objects
- Highlight some other deployed VRE's that would benefit to the sorts of improvements you suggest. You could include Wakari.io and also the tmpnb thing in Nature magazine: <http://www.nature.com/news/ipython-interactive-demo-7.21492>

## 1.2 Relation to the Work Programme

♠EC Commentary: 1-2 pages; Eugenia will help there♠

♠EC Commentary: *Indicate the work programme topic to which your proposal relates, and explain how your proposal addresses the specific challenge and scope of that topic, as set out in the work programme.*♠

♠EC Commentary:

This is a synthesis of Antonios Barbas' slides describing the Call 3  
Topic 9-2015 EINFRA-9: e-Infrastructure for Virtual Research Environment

See file:../Documentation/VirtualEnvironmentsWorkProgramme2014-2015.ppt

\*\* Suggested EU contribution per proposal: 2 to 8 M euros ; Total budget: 42 Meuros

\*\* Dates: 14/01/2015

- H2020-EINFRA-2014-1 15/04/2014
- H2020-EINFRA-2014-2 02/09/2014
- H2020-EINFRA-2015-1 14/01/2015(tbc)

\*\* European contacts: Antonios Barbas  
See file:Documentation/Contacts.docx

\*\* Definition:

- Groups of researchers, typically widely dispersed, who are working together
- through ubiquitous, trusted and easy access to services for scientific data, computing and networking
- in a collaborative, virtual environment:
  - > the e-Infrastructures

\*\* Characteristics:

- Address the needs of specific scientific communities { in support of e-Science;
- Have users from both academia and industry;
- Involve bottom-up research and develop user-oriented services;
- Are based on e-infrastructures

\*\* Specific challenge:

- Capacity building in interdisciplinary research
- through community-led development and deployment of service-driven digital environments
- for large-scale cross-disciplinary research collaboration and data interoperability

\*\* Expected impact:

- More effective collaboration between researchers and increased take-up of collaborative research by new disciplines;
- Easier discovery, access and re-use of data, resulting in higher productivity of researchers;
- Accelerate innovation via access to integrated digital research resources across disciplines;

\*\*\* Scope: Proposals are expected to

Notations: [X]: easy to argue; [?]: we have some lead, but that will take some arguing

- [?] Integrate resources across all layers of the e-infrastructure (networking, computing, data, software, user interfaces) to foster cross-disciplinary data interoperability
- [?] Build on requirements from real use cases, i.e. integrate heterogeneous data from multiple sources and re-use tools and services from existing infrastructures
- [X] Target any area of Science and Technology, especially

- interdisciplinary ones, including ICT, mathematics, web science and social sciences and humanities
- [X] Use standardised building blocks and workflows, well-documented interfaces and interoperable components;
- [?] Define semantics, ontologies and metadata to enable data citation and promote data sharing, as to ensure interoperability;
- [X] Target easy-to-use functionalities; and indicate the number of researchers they target as potential users;
- \*\* Specific conditions for the Call on e-Infrastructures:
  - [X?] Proposals should be structured around Networking, Service and Joint Research Activities
  - [X] The Software to be developed needs to be open source
  - [ ] A Data Management Plan to be developed enabling data preservation, on-line discoverability, authorisation and re-use of data
  - [X] Clear Metrics (KPIs) to be proposed and used;
  - [?] Open Access to Publications resulting from the project;
  - [X] Usefulness of services to the end user community and financial sustainability to be ensured;
- \*\* Where should the emphasis be?
  - [?] Services
  - [X] Thinking innovation
    - With both suppliers or users
  - [X] Mainstreaming skills development
  - [ ] Integration between data and computing
  - [X] Business plans for financial sustainability
    - ...and partnerships with the private sector
  - [ ] Supporting policies
  - [X] open data and software
  - [X] Sharing basic operations services and building blocks
  - [X] Monitoring performance (KPIs)
- \*\* Simplified funding model
  - Up to 100% for Research and Innovation
    - Flat 25% rate for indirect costs (overhead?)



### 1.3 Concept and Approach

♠**EC Commentary:** 5-8 pages♠ ♠**EC Commentary:** – Describe and explain the overall concept underpinning the project. Describe the main ideas, models or assumptions involved. Identify any trans-disciplinary considerations; – Describe and explain the overall approach and methodology, distinguishing, as appropriate, activities indicated in the relevant section of the work programme, e.g. Networking Activities, Service Activities and Joint Research Activities, as detailed in the Part E of the Specific features for Research Infrastructures of the Horizon 2020 European Research Infrastructures (including e-Infrastructures) Work Programme 2014- 2015; – Describe how the Networking Activities will foster a culture of co-operation between the participants and other relevant stakeholders. – Describe how the Service activities will offer access to state-of-the-art infrastructures, high quality services, and will enable users to conduct excellent research. – Describe how the Joint Research Activities will contribute to quantitative and qualitative improvements of the services provided by the infrastructures. – As per Part E of the Work Programme, where relevant, describe how the project will share and use existing basic operations services (e.g. authorisation and accounting systems, service registry, etc.) with other e-infrastructure providers and justify why such services should be (re)developed if they already exist in other e-infrastructures. Describe how the developed services will be discoverable on-line. – Where relevant, describe how sex and/or gender analysis is taken into account in the project’s content.♠

#### 1.3.1 Linked research and innovation activities

♠**EC Commentary:** Describe any national or international research and innovation activities which will be linked with the project, especially where the outputs from these will feed into the project;♠

♠**TO DO:** For each item below, write a paragraph describing the project and one describing how it connects with this proposal♠

#### DFG Priority Project SPP 1489 [computer.algebra.de](http://computer.algebra.de)

The SPP1489 “Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory” is a nationwide Priority Project of the German Research Council DFG which commenced in July 2010 and will end in June 2016. The focus of the programme is on the interactions between computer algebra and algebraic geometry, number theory, and group theory. It combines expertise at all levels of research in computer algebra, be it the design of algorithms, the implementation of algorithms, the application of algorithms, or the creation of mathematical databases. The goal of SPP1489 is to considerably further the algorithmic and experimental methods in the afore mentioned disciplines, to combine the different methods across boundaries between the disciplines, and to apply them to central questions in theory and praxis. A fundamental concern of the programme is the further development of open source computer algebra systems with origins in Germany, which in the framework of different projects will be crosslinked on different levels. Of particular interest are interactions with application areas inside and outside of mathematics such as system- and control theory, coding theory, cryptography, CAD, algebraic combinatorics, and algebraic statistics as well as hybrid methods which combine numerical and symbolic approaches.

♠**WD [WRITE HERE:** One paragraph description of how this relates to this project]♠

#### IPython/Jupyter grant from the Alfred P. Sloan foundation <http://ipython.org/sloan-grant.html>

♠**IPython [WRITE HERE:** Proofread description of the Sloan grant and link to this project]♠

The IPython project received a \$1.15M grant from the Alfred P. Sloan foundation that is supporting IPython development for two years (1/1/2013-12/31/2014), in particular at the University of California, Berkeley and California Polytechnic State University, San Luis Obispo. This grant enabled the project to focus on developing the IPython Notebook as a general tool for scientific and technical computing that is open, collaborative and reproducible. This goes a long way toward Aim ♠**TO DO:** ... and ...♠ of **MathVRE**, especially given the current rapid evolution of IPython toward its language agnostic avatar Jupyter.

**MathVRE** will build on the outcome of the Sloan grant, and further develop the critical IPython/Jupyter component in close collaboration with the IPython/Jupyter team. In particular, we plan to hire some of the European developers that are currently funded by the Sloan grant to work in California and wish to later return to Europe.



**Sage-Combinat grant** ♠NT [WRITE HERE: ...]♠

**Logilab: simulagora, cubicweb, ...** ♠Logilab [WRITE HERE: One paragraph description of simulagora, cubicweb, ...]♠ ♠Logilab [WRITE HERE: How does it relate to this project]♠

**SageMathCloud** <https://cloud.sagemath.com/>

♠NT/SL/WS [WRITE HERE: Proofread section about SMC]♠

This section was cowritten with William Stein, head of the SageMathCloud project.

SageMathCloud provides a collaborative online environment for students, teachers and researchers to interact with Sage and with each other. It has Sage and IPython worksheets, powerful  $\text{\LaTeX}$  editing features and a full Linux computer, all accessible from a standard web browser. Its main design feature is to enable and promote collaboration between groups of users. It is for example a natural place to host a course, allowing teachers to collaborate with their students using modern tools like Sage and  $\text{\LaTeX}$ , with facilities for real-time communication through chat, video, and shared editing of documents, programs and worksheets; course material can be provided as worksheets, assignments can be distributed, collected, and returned as well. Launched in 2013, SageMathCloud presently hosts over 100,000 projects and 10,000 weekly active users. This fast adoption by a wide variety of users demonstrates the relevance and the long term impact this kind of collaborative environments can have.

Technically speaking, SageMathCloud is a specific cloud-based Virtual Research and Teaching Environment for mathematics, based on open source technology. It's currently developed and deployed at the University of Washington at Seattle, with a business plan in the work for commercial support for massive on line courses, subsidizing free access for all other academic usage some further Sage development.

In comparison **MathVRE** focuses on open source building blocks and architecture to easily setup and deploy custom Virtual Research Environments. On the one hand, SageMathCloud will serve as prototype for **MathVRE**, paving the way and showcasing important features from the users perspective. On the other hand, basically each and every task undertaken in **MathVRE** will benefit back SageMathCloud.

## FLINT grant?

**LMFDB grant** The L-functions and Modular Forms Database (LMFDB) project originated at a meeting at The American Institute for Mathematics (AIM) in 2007. L-functions are ubiquitous in number theory, and have applications to mathematical physics and cryptography. The simplest example of an L-functions is the Riemann zeta function. Two of the seven Clay Mathematics Million Dollar Millennium Problems deal with properties of these functions, namely the Riemann Hypothesis and the Birch and Swinnerton-Dyer Conjecture, that were conjectured following computational exploration. As well as providing a central repository of data as a resource for researchers, through its website [www.lmfdb.org](http://www.lmfdb.org), the LMFDB provides a modern handbook, including tables, formulas, links and references, concerning particular specific L-functions and their sources. Between 2008 and 2012 the LMFDB was funded through a US National Science Foundation (NSF) Focussed Research Grant (FRG) of around \$1M. Since 2013, the funding of the LMFDB has passed to Europe through a six year £2.2M Programme Grant from the UK Engineering and Physical Sciences Research Council (EPSRC), held at the universities of Warwick and Bristol, with Professor John Cremona (Warwick) as its Principal Investigator (see [http://www2.warwick.ac.uk/fac/sci/maths/people/staff/john\\_cremona/lmf](http://www2.warwick.ac.uk/fac/sci/maths/people/staff/john_cremona/lmf)). This grant supports six three-year postdoctoral research fellows, mathematical researchers who work on the mathematical aspects of the project full-time, biannual workshops, equipment and a portion of the investigators' own time.

Almost all contributors to the LMFDB project, including those directly supported by the EPSRC grant and the larger world-wide team of 30-50 contributors of data and code, are pure mathematicians. Most of these have good computational skills, but are not professional programmers or software developers. The LMFDB has a great need to broaden the support it can call upon from software developers, to enhance the project in several ways, including the computation of number-theoretic data but more specifically in supporting the database management and website user interface, in order to make the data more accessible and useful to others. The codebase of the LMFDB project is entirely open source and hosted at github (<https://github.com/LMFDB/lmfdb>), written in python with specialist modules such as flask and pymongo to manage the website and database interface, and Sage for higher-level mathematical computations. The LMFDB project would therefore benefit greatly from collaboration with **MathVRE** as it would connect the



project with a pool of experts. Joint workshops between the LMFDB and **MathVRE** will stimulate and develop such collaboration: the LMFDB places great importance on its workshops, which are small gatherings of around 30 invited participants who work throughout one week on certain specific aspects of the project, coming together in plenary sessions to make decisions, plan and collectively approve of proposed developments. As a leading example of the use of databases in mathematical research, the LMFDB will provide **MathVRE** with a real large-scale prototype around which to develop new ideas about the design and implementation of such databases and their associated software. The feasibility of such collaboration was successfully tried at a workshop at the ICMS in Edinburgh in January 2013 on “Online databases: from L-functions to combinatorics”, sponsored by the NSF, AIM and the ICMS.

**Findstat?**

**Kwarc group**

## 1.4 Ambition

♠EC Commentary: 1-2 pages♠

♠EC Commentary: – Describe the advance your proposal would provide beyond the state-of-the-art, and the extent the proposed work is ambitious. Your answer could refer to the ground-breaking nature of the objectives, concepts involved, issues and problems to be addressed, and approaches and methods to be used.

– Describe the innovation potential which the proposal represents. Where relevant, refer to products and services already available, e.g. in existing e-Infrastructures.♠

## 2 Impact

♠**TO DO:** Orsay's grant services will help here in December♠

### 2.1 Expected Impacts

♠**EC Commentary:** *Please be specific, and provide only information that applies to the proposal and its objectives. Wherever possible, use quantified indicators and targets.*

*Describe how your project will contribute to:*

- the expected impacts set out in the work programme, under the relevant topic (including key performance indicators/metrics for monitoring results and impacts);*
- improving innovation capacity and the integration of new knowledge (strengthening the competitiveness and growth of companies by developing innovations meeting the needs of European and global markets; and, where relevant, by delivering such innovations to the markets;*
- any other environmental and socially important impacts (if not already covered above).*

*Describe any barriers/obstacles, and any framework conditions (such as regulation and standards), that may determine whether and to what extent the expected impacts will be achieved. (This should not include any risk factors concerning implementation, as covered in section 3.2.)♠*

## 2.2 Measures to Maximise Impact

### 2.2.1 Dissemination and Exploitation of Results

♠**EC Commentary:** – *Provide a draft 'plan for the dissemination and exploitation of the project's results'. The plan, which should be proportionate to the scale of the project, should contain measures to be implemented both during and after the project.*

*Dissemination and exploitation measures should address the full range of potential users and uses including research, commercial, investment, social, environmental, policy making, setting standards, skills and educational training.*

*The approach to innovation should be as comprehensive as possible, and must be tailored to the specific technical, market and organisational issues to be addressed*

– *Explain how the proposed measures will help to achieve the expected impact of the project . Provide a draft business plan for financial sustainability as stated in the Part E of the Specific features for Research Infrastructures of the Horizon 2020 European Research Infrastructures (including e-Infrastructures) Work Programme 2014-2015.*

– *Where relevant, include information on how the participants will manage the research data generated and/or collected during the project, in particular addressing the following issues: What types of data will the project generate/-collect? What standards will be used? How will this data be exploited and/or shared/made accessible for verification and re-use (If data cannot be made available, explain why)? How will this data be curated and preserved?*

– *Include information about any open source software used or developed by the project.*

*You will need an appropriate consortium agreement to manage (amongst other things) the ownership and access to key knowledge (IPR, data etc.). Where relevant, these will allow you, collectively and individually, to pursue market opportunities arising from the project's results.*

*The appropriate structure of the consortium to support exploitation is addressed in section 3.3.*

– *Outline the strategy for knowledge management and protection. Include measures to provide open access (free on-line access, such as the “green” or “gold” model) to peer-reviewed scientific publications which might result from the project.*

*Open access publishing (also called 'gold' open access) means that an article is immediately provided in open access mode by the scientific publisher. The associated costs are usually shifted away from readers, and instead (for example) to the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research.*

*Self-archiving (also called “green” open access) means that the published article or the final peer-reviewed manuscript is archived by the researcher - or a representative - in an online repository before, after or alongside its publication. Access to this article is often - but not necessarily - delayed (“embargo period”), as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download/view fees during an exclusivity period.♠*

**Long term sustainability** By design (Objective 1), the VRE's promoted by **MathVRE** will consist of a thin layer on top of an ecosystem of components. Hence, the long term sustainability of those VRE is guaranteed by the sustainability of the ecosystem of components, that is by Objective 5.

### 2.2.2 Communication activities

♠**EC Commentary:** *Describe the proposed communication measures for promoting the project and its findings during the period of the grant. Where appropriate these measures should include social media and public events with user participation. Measures should be proportionate to the scale of the project, with clear objectives. They should be tailored to the needs of various audiences, including groups beyond the project's own community. Where relevant, include measures for public/societal engagement on issues related to the project.*♠

### 3 Implementation

♠**TO DO:** *Typical granularity: 5-8 work packages with 3-5 tasks and one deliverable per task; 10 milestones*♠

#### 3.1 Work Plan — Work packages, deliverables and milestones

♠**EC Commentary:** *Please provide the following:*

- *brief presentation of the overall structure of the work plan;*
- *timing of the different work packages and their components (Gantt chart or similar);*
- *detailed work description, i.e.:*
  - *a description of each work package (table 3.1a);*
  - *a list of work packages (table 3.1b);*
  - *a list of major deliverables (table 3.1c);*
- *graphical presentation of the components showing how they inter-relate (Pert chart or similar).*

♠

##### Overall Structure of the Work Plan

The work plan is broken down into XX workpackages as shown in Figure ??: WP2 deals with ... In addition, there is one management work package (WP1) and one general dissemination work package (??). The Gantt chart on Page 21 illustrates the timeline for the various tasks for these work packages, including inter-task dependencies.

##### How the Work Packages will Achieve the Project Objectives

♠**ALL [WRITE HERE:** This needs to explain that we're actually going to meet the objectives. Needs to be done after objectives and WPs.]♠

The project objectives (Section 1.1, page 2) and the corresponding work packages that contribute to achieving those objectives are:

Objective	Purpose	WPs
Objective 1	XX	WPX

**Work Programme for Objective 1:** Objective 1 is covered by WPX, which will ...



## Work package list

Work package No	Work package title	Lead partic. no.	Lead short name	Person months	Start month	End month
WP1	Project Management	1	UPS		1	60
WP2	Community Building and Engagement	1	UPS			
WP3	Component Architecture	1	UPS			
WP4	User Interfaces	1	UPS			
WP5	HPC and massively parallel components	1	UPS			
WP6	Next generation mathematical Data/Knowledge/Software-bases	9	USTAN			
WP7	Development Models for an Academic Free Software Ecosystem					
WP8	Social Aspects	6	UO			
WP9	Dissemination, Exploitation and Communication	9	USTAN			
<b>Total</b>				<b>XXX</b>		



<b>List of Deliverables</b>
-----------------------------

Del. no.	Deliverable name	WP no.	Lead	Type	Dissemi- nation level	Delivery date
-------------	------------------	-----------	------	------	-----------------------------	------------------

## List of milestones

Milestone number	Milestone name	Related work package(s)	Estimated date	Means of verification (deliverables shown here + success criteria below)
MS1 MS3	Completed initial requirements analysis.	WPX WPX	1	??.

Milestone	Success Criteria	Contributes to Objective(s)
MS1 MS3	Completed requirements analysis (Deliverable ??). XX	<b>1, 3.</b> <b>XX</b>

♠**EC Commentary:** *KEY Estimated date Measured in months from the project start date (month 1) Means of verification Show how you will confirm that the milestone has been attained. Refer to indicators if appropriate. For example: a laboratory prototype that is ‘up and running’; software released and validated by a user group; field survey complete and data quality validated. ♠*

## Work package description (WP1)

Work package number	WP1		Start date or starting event:			Month 1	
Work package title	Project Management						
Participant number	1	9					
Participant short name	UPS	USTAN					
Person-months per participant:	48	48					

**Objectives:** Objectives: The objectives of WP1 are to undertake all project management activities, including:

- monitoring the overall progress of the project and the use of resources;
- ensuring the timely production of deliverables and other project outputs;
- reporting to the European Commission on financial matters;
- preparing for and attending the annual project review meetings; and
- managing the project Advisory Board.

### **Description of work:**

This workpackage will perform all the activities related to monitoring of progress towards the project milestones shown on Page 3.1 and the deliverables listed on Page 3.1, assuring the quality of the deliverables, ensuring the collation and distribution of the required reports, questionnaires and deliverables including the annual reports to the European Commission, arranging project management meetings, tracking the project budget in terms of expenditure and person-months, obtaining financial certificates as required, convening project management meetings, ensuring that important project documents such as the project contract and the consortium agreement are properly maintained and amended as necessary, ensuring that contractual details are complied with, monitoring compliance with the grant agreement, preparing for the annual review meetings, and reviewing research results against the aims and objectives of the project. It also involves managing and supporting the project Advisory Board, including supporting attendance at project meetings, convening Advisory Board meetings, and obtaining feedback on the project direction and results.

**Deliverables:****Deliverable 1.1** *(Month 1) Create tickets for all relevant tasks / deliverables***♣TO DO:** *Rewrite everything below using the WPDeliverable macro♣*

- D1.1Month 1): Internal and external mailing lists.
- D1.3Month 1): Internal software repository.♣**TO DO:** *Needed?*♣
- D1.3Month 1): Internal software repository.♣**TO DO:** *Needed?*♣
- D1.4Month 12): Project Periodic Report (first year).
- D1.5Month 24): Project Periodic Report (second year).
- D1.6Month 36): Project Periodic Report (third year).
- D1.7Month 48): Project Periodic Report (fourth year).
- D1.8Month 48): Project Final Report

## Work package description (WP2)

Work package number	WP2	Start date or starting event:					Month 1
Work package title	Community Building and Engagement						
Participant number	9						
Participant short name	USTAN						
Person-months per participant:	1						

**Objectives:** The objective of WP2 is to further develop the community at the European scale, foster cross teams collaborations, spread the expertise, and engage the greater community to participate to the definition of the needs, and the implementation and use of the produced solutions.

### **Description of work:**

We will organize regular open workshops (e.g. Sage Days, Pari Days, summer schools, etc.); some of them will be focused on development and coding sprints, and others on training.

This work package will also provide general travel budget to fund short to long term visits between the participants, to collaborate on specific features. A typical such visit would bring together an IPython developer with a GAP developer for a couple of days to implement a first prototype of notebook interface to GAP.

This work package will complement and lean on a parallel COST network whose role is to build and animate the greater community.

### **Deliverables:**

- ?? (Month 12): Report on community needs
- Workshop 1 ...
- Workshop 2 ...
- Workshop 3 ...
- ♠**TO DO: make a list**♠

## Work package description (WP3)

Work package number	WP3		Start date or starting event:			Month 1	
Work package title	Component Architecture						
Participant number	1	3	9				
Participant short name	UPS	UVSQ	USTAN				
Person-months per participant:	24	1	1				

**Objectives:** The objective of this work package is to develop and demonstrate a set of API's enabling components such as database interfaces, computational modules, separate systems such as GAP or Sage to be flexibly combined and run smoothly across a wide range of environments (cloud, local, server, ...).

### Description of work:

#### Task 3.1 *Portability*

*In order to achieve maximum availability and accessibility, mathematical software must be developed and tested for a wide range of computer architectures and operating systems. However most of open source development happens in POSIX environments (usually Linux or OSX), and almost exclusively on x86 platforms. The vast majority of the developers of mathematical software does not have the expertise, nor the access to appropriate hardware and software, to insure appropriate testing and porting of components. The best incarnation of this issue is the involved installation procedure for Sage on Windows, a major adoption barrier and common source of complaints by end-user.*

*In this task we will address the common needs of the community in terms of portability layers, building and testing infrastructure.*

- *Best practices adopted by the larger open source community will be investigated and leveraged, and existing expertise will be shared between the component developers.*
- *Windows being largely dominant in the desktop/laptop market, a specific focus will be placed on the port of Sage, and therefore all the components included in its distribution (in particular Pari, GAP, Singular, Linbox) to this platform (D 3.2, D 3.3).*
- *The deployment of a common infrastructure for multi-platform continuous integration (testing, building and distribution) will be addressed.*

#### Task 3.2 *Interfaces between systems*

*In this task we will investigate patterns to share data, ontologies, and semantics across computational systems, possibly connected remotely. We will leverage the well established semantics used in mathematics (categories, type systems, ... ) to give powerful abstractions on computational objects.*

*Through well defined APIs, we will enable discovery of subsystems, functionality, documentation and computational resources. The user interfaces shall be enabled to automatically choose the best available algorithms and resources to perform a required computation, as well as clearly and intuitively present the available choices to the expert user. As a first concrete test bed, we will consider the Sage interface to GAP, or more precisely libGAP (Deliverable 3.4). Like most Sage interfaces, this uses the now classical handle design pattern, whereby one can manipulate from Sage an object created and stored in GAP, through a handle (a.k.a. remote objects). By introspection, one can discover available GAP functions on this object and access the related GAP documentation. In a few cases, for example when the object is a group, one can further wrap the handle with an adapter that makes it behave like a native Sage group. By mapping GAP's categories to Sage's categories (as a rough approximation, both can be thought of as hierarchies of abstract classes), we will:*

- Implement a modular infrastructure for adapters, based on the category hierarchy, in order to let the implementation of adapters scale to a large variety of objects.
- Refactor the existing adapters, using this infrastructure to generalize their features. This step by itself will provide adapters for larger categories like semigroups or monoids.
- Merge the adapters into the handles, so that a handle to a GAP group will automatically behave like a native Sage group.

A specific challenge will be performance; indeed low level method adapters, e.g. for arithmetic, need to be compiled when most of the interface infrastructure is dynamic by nature.

♠**TO DO:** Other deliverables: generalization to other interfaces, factorization of the adapter code through standardized API's, semantic-enabled remote objects in SCSCP, reports, ...♠

### **Task 3.3 Modularization and packaging**

In this task we will investigate best practices for composing, sharing and interfacing computational components and data for connected mathematical systems.

We will start with a comparative study of the practices adopted in various open source projects, both inside and outside of this project. This will include reviewing non-mathematical systems, e.g.: operating systems, platforms, web frameworks, cloud and HPC infrastructures.

We will address the current shortcomings to promote a new generation of mathematical software that is capable of scaling to large code bases, large datasets, and massively distributed infrastructures. This task also needs to consider the results of work package WP8n social issues regarding distributed development, community management, acknowledging contributions, etc.

As an example, Sage has a long history of integrating and distributing large mathematical libraries/software as a whole, with relatively few attention given to defining and exposing interfaces. Component re-usability is not a main focus for the Sage community, at the same time the non-standard and relatively underused package system discourages writing and maintaining autonomous libraries. These factors have contributed to make the Sage distribution what is usually described as a “monolith” (Sage library code alone, not counting included libraries, makes up for 1.5M lines of code and documentation), hard to distribute, to maintain, to port, and to develop with.

On the opposite side, GAP has been distributing community-developed “GAP packages” for a long time, but faces now fragmentation issues, at the code and at the community level. The rudimentary package system adds more technical difficulties to GAP's development model.

Both models reach the limits of their scalability, and a synthesis is very much needed.

### **Task 3.4 Deployment and distribution**

♠**TO DO:** NT: what did you have in mind?♠

### **Task 3.5 Component architecture for High Performance Computing and Parallelism**

As in all other areas of science, properly supporting massively parallel architecture is a major challenge. Many of the computational components have already gone a long way in this direction, and further work will happen there within WorkPackage T5.1.

In this task we will investigate and implement parallelism-friendly ways of combining components together, so that calling components can benefit from the parallelism features of called components, with self-adaptation to the environment and cooperative sharing of resources. ♠**TO DO:** details+deliverables♠

### **Task 3.6 Improving the development workflow, in Sage and elsewhere.**

♠**TO DO:** ♠

## Deliverables:

**Deliverable 3.2** (Month 12) one-click install Sage distribution for Windows with Cygwin 32bits

**Deliverable 3.3** (Month 24) one-click install Sage distribution for Windows with Cygwin 64bits

**Deliverable 3.4** (Month 12) Semantic-aware Sage interface to GAP.

♠**TO DO:** *Make all those deliverables one-liners, and move the material to the task descriptions*♠

- ?? Make sure that Sage and therefore all the components it depends on (including GAP, Linbox, Pari, Singular, ...) have standard packages in the main Linux distributions: Debian/Ubuntu, Redhat, Gentoo, ...

♠**TO DO:** *Get feedback from our experts, and make this precise; what can we actually promise to achieve? how much work is this? Do we have personnel for this? There is strong expertise in Logilab with a Debian developer working there; he could advise someone on this. Logilab is interested in this because it's meeting similar issues with some of its clients software like Salomé.*♠

- ?? (Month 12): Creation, deployment, and distribution of preconfigured virtual machines (and/or Docker images) for Pari, Sage, ... as a cloud service, in particular within the StratusLab infrastructure. This includes build bots and test bots for continuous integration over a variety of operating systems.

- ?? Modularization of the Sage distribution

Separation of the different components of Sage (communication with third-party softwares, build system, Sage native code). This is a prerequisite for easier packaging and integration in standard Linux distributions and lmonade, native integration within the IPython notebook and other interfaces (larchenv, Spyder, ...) and collaboration with sister projects.

- ?? Add support for the SCSCP interface protocol to all relevant components (e.g. Sage, ...). ♠**SL/AK [WRITE HERE: Brief description of what SCSCP is, reference to previous grant, relevance to the goals of this grant; maybe this should go in the work package description]**♠

SCSCP (Symbolic Computation Software Composability Protocol) is a remote procedure call protocol by which a computer algebra system (CAS) may offer services to a variety of possible clients, including e.g. another CAS running on the same computer system or remotely; another instance of the same CAS (in a parallel computing context); a simplistic SCSCP client (e.g. C/C++/Python/etc. program) with a minimal SCSCP support needed for a particular application; a Web server which passes on the same services as Web services, etc. A distinctive feature of the protocol is that both instructions and data are represented in the OpenMath format (<http://www.openmath.org/>; previously supported by the EU JEM Thematic Network; EU project 24969 "ESPRIT" and other projects); moreover, OpenMath support is not limited by existing official OpenMath content dictionaries - private encodings may be easily embedded into SCSCP messages.

SCSCP has been developed in the EU FP6 project 26133 "SCIENCE – Symbolic Computation Infrastructure for Europe" (TODO: fix and insert URL) and by now is supported by a number of computer algebra systems, including GAP, Macaulay2, Maple, TRIP and others. To facilitate SCSCP implementations, there are also APIs for Java, C and C++, and a simple Python SCSCP client (for a full overview, see SCIENCE website).

TODO: more on the relevance to the goals of this grant. This is useful to exchange information between systems for problems that can not be solved within any single system; for storing and retrieving information (in databases) immediately into the CAS session; for organising distributed computations.

- Some IPython/Jupyter deliverables here. ♠**TO DO:** *review what it can already do in term of choice of computational resource and storage back-end.*♠ ♠**TO DO:** *cross link to Task 4.7*♠

- Contribution by Kaiserslautern: libSingular, pySingular?, GAP-Singular, Singular-Sage.

Moving code from Sage into Singular when relevant

- ?? (Month ...) Configure the components of Sage's distribution (e.g. Atlas, Linbox, GAP, Singular, ...) to be systematically HPC-enabled, and make sure that Sage's calls to such components indeed enable HPC.

MathVRE • Transparent integration of IPython capabilities for cluster computing.

- Implementation of a transparent abstraction over mpi.

- Develop or integrate existing solutions for MapReduce operations over big data.



Raw material:

## Component Architecture

-----

Recomputation connection belongs here?

Collaboration with unreliable (or restricted!) networking connections (peer-to-peer, opportunistic syncing, 3rd world). This is technically interesting, and gets in support for non-networked working. Not sure if it belongs here or not.

- Security concerns

Goal: Fostering collaborations/integration between components in an open source ecosystem

=====

- How to make systems "cooperate" rather than "predate each other".
- E.g. reduce the version issues
- Foster collaboration with upstream libraries by sharing the development and maintenance of the interfaces, typically as standalone upstream Python bindings (e.g. py-Singular).
- How to make it easy to develop simultaneously two interdependent components (e.g. Sage+Singular)
- Foster communication
- Social aspect:  
Credit, Citations, Recognition, Funding

## Documentation system

=====

In which package?

## Improvements to Sphinx

Sage heavily customizes the Sphinx documentation system, hacking deep in it in some cases, with quite some duplication in some cases. Refactor the whole thing, generalizing and contributing back upstream as much as possible (e.g. parallel compilation).

## Work package description (WP4)

Work package number	WP4		Start date or starting event:			Month 1	
Work package title	User Interfaces						
Participant number	1	2	9	11	5		
Participant short name	UPS	Logilab	USTAN	UK	UB		
Person-months per participant:	1	1	1	1	1		

**Objectives:** The objective of this work package is to provide a modern, robust, and flexible user interface for computation, supporting real-time sharing, integration with collaborative problem-solving, multilingual documents, paper writing and publication, links to databases, etc.

### Description of work:

♠**TO DO:** *What is a notebook interface*♠

♠**IPython** [WRITE HERE: improve this draft presentation of IPython]♠

IPython is a leading notebook interface in the world of interactive computations, and use massively by biologists, physicists, ♠**TO DO:** *and outside academia!*♠. Originally tailored for Python, it has been language agnostic, and can communicate through a standardized interface to various computation kernels ♠**TO DO:** *cite a few*♠. It can transparently run kernels locally or remotely (e.g. on the cloud), and has built in support for parallel computing and HPC.

### Task 4.7 Uniform notebook interface for all interactive components

*In this task, we will implement Jupyter interfaces for the interactive computation components of **MathVRE**, including GAP, Pari, Sage, and Singular. A first release D 4.5 will focus on basic functionality, and a second release D 4.6 will cover advanced features like 3D graphics or transparent documentation browsing (as live worksheets whenever relevant).*

*Sage itself will require a specific treatment as it already has a notebook interface. Its development started about at the same time as the IPython notebook, with similar target features but a different agenda: the Sage notebook had to be available very quickly to solve pressing needs of the Sage community; instead the IPython notebook was to take its time and build robust foundations from the ground up. The two projects have exchanged a lot, and the IPython notebook, which benefits from a much larger user base and thus developer pool, has mostly caught up with the Sage notebook in terms of functionality. It's thus time for the Sage community to outsource this key but non disciplinary component and phase out the Sage notebook in favor of the IPython notebook.*

*The Sage and IPython convergence D 4.7 will require:*

- Robust migration path and tools for Sage worksheets,
- Support for math, 2D, and interactive 3D output.,
- Import (and export?) of ReST documents, with full support for Sage's specific roles (math, ...),
- Support for remote Sage kernel, typically on the cloud, or running with a different Python version (Sage as a library),
- A migration path for interactive widgets implemented with Sage's @interact functionality.

*Joint meetings and visits between the developers of Jupyter and of the computing components will be a key asset for this task.*

♠**TO DO:** *convergence with TeXmacs?*♠

#### Task 4.8 Notebook interface usability

In this task, we will further improve Jupyter as a uniform user interface for interactive computation and database query.

♠IPython [WRITE HERE: Split off the work into an appropriate list of deliverables]♠

- Collaborative and shared worksheets,
- Multilingual notebooks?
- Improved 2D/3D graphics: maybe architecture for integrating VPython, vispy, ...? 3D visualization of algebraic surfaces (in collaboration with Singular)
- Native folding support for sections and the like, with mouse and keyboard commands.
- Version control
- Reproducibility: ?? Support for tested notebooks. The writer specifies the expected outputs, e.g. in text format, and can check at any point that the full execution of the notebook yields exactly the expected output, as can be done with e.g. ReST files in Sage: `sage -t notebook.rst`
- ?? (Month ???) Heavyweight (e.g. QT based and not web based) user interface with docking support in the style of Spyder <https://code.google.com/p/spyderlib/>. Can possibly be implemented by extending the QT IPython console, or by letting Spyder use the IPython protocol (♠TO DO: If that's not yet the case♠).

#### Task 4.9 Dynamic documentation and exploration system

Introspection has become a critical tool in interactive computation, allowing user to explore on the fly the properties and capabilities of the objects under manipulation. This becomes particularly acute in systems like Sage where large parts of the class hierarchy is built dynamically, and static documentation builders like Sphinx cannot anymore render all the available information.

In this task, we will investigate how to further enhance the user experience. This will include:

- On the fly generation of Javadoc style documentation, through introspection, allowing e.g. the exploration of the class hierarchy, available methods, etc.
- ♠Logilab [WRITE HERE: Inclusion of database queries and views]♠
- ?? (Month 36) Exploratory support for semantic-aware interactive widgets providing views on objects represented and or in databases

Preliminary steps are demonstrated in the Larch Environment project (see demo vidéo on <http://www.larchenvironment.com/>) and sage-explorer.

Ultimate goal: automatically generated LMFDB-style interfaces.

Whenever possible, those features will be implemented generically for any computation kernel by extending the Jupyter protocol with introspection and documentation queries.

.

#### Task 4.10 Structured documents

Support for writing interactive structured documents, and in particular papers, books, experimentation log books and reports, presentations, course notes, etc, with the following features:

- Static printed/PDF/HTML version and interactive version.  
Achieved by either importing or exporting document files in some standard format (LaTeX, ReST, Markdown, ...).

- *Tests (see above).*
- *Collaborative edition.*
- *Version control.*

♠**TO DO:** *include here everything about this topic in Needs.rst*♠

♠**TO DO:** *Wherever relevant, create tickets with details, and refer to them here.*♠

#### **Deliverables:**

**Deliverable 4.5** *(Month 12) Basic Jupyter interface for GAP, Pari, Sage, Singular*

**Deliverable 4.6** *(Month 12) Full featured Jupyter interface for GAP, Pari, Singular*

**Deliverable 4.7** *(Month 12) Sage notebook / IPython notebook convergence*

About the availability of people to hire, I have a full-time, experienced developer whose contract runs out in fall 2015, he would be ideal for the project. I also have a doctoral student who needs employment after the MathSearch project (until 10/2015) runs out. So I do have people who would directly be available.

Michael

=====8<-----

Task 4.10. Structured Documents (12 PM total, 3 PM per deliverable)

-> This existing task we could just take over based on our MathHub.info system, which would need to be adapted to the task.

Deliverables:

D1: Active Documents based on sTeX

D2: Distributed, Collaborative, Versioned Editing of Active Documents

in MathHub.info

D3: Notebook Import into MathHub.info (interactive display)

D4: in-place computation in active documents (context/computation).

Comments:

MathHub.info is a portal for reading and interacting with "active documents"

(i.e. documents that have an additional semantic layer that supports semantic services like

- definition lookup, type-inference, unit conversion, ...)

Notebooks are essentially "programs with documentation", whereas active documents are

documents with a semantic knowledge layer. Regular publications are an important

boundary case: Active Documents look like papers, but are web-standards compatible and interactive.

sTeX is a semantic variant of LaTeX that we can transform into OMDoc/MMT, which is the native knowledge representation format for active documents and machine-actionable knowledge about math and symbolic programs.

=====8<-----

Task K-4.11 Math Search Engine (10 PM total; 2 each for D1/2, 3 each for D3/4)

D1: Full-text search (formulae + Keywords) over LaTeX-based documents (e.g. arXiv subset)

D2: Full-text search (F+K) over Notebooks (in the format determined in task 4.7)

D3: Formula search in CAS programs and Software Modules

D4: Search from Notebooks/Active Documents (for local context to inform search)

Comments:

We already have a search engine, therefore we only need to build harvesters for D1/2;

D3/4 are more speculative.

I am going to get some sushi now, I have 2 or 3 tasks for WP6 in mind, but not written up.

## Work package description (WP5)

Work package number	WP5		Start date or starting event:			Month 1	
Work package title	HPC and massively parallel components						
Participant number	1	2	9	11	5	4	
Participant short name	UPS	Logilab	USTAN	UK	UB	UJF	
Person-months per participant:	1	12	1	1	1	12	

**Objectives:** The objective of this work package is to improve the performance of the computational components of **MathVRE**, in particular on massively parallel architectures. This includes notably:

- Fine grained High Performance Computing on many-cores architectures.
- Coarse grained or embarrassingly parallel computing on grids or on the cloud.
- Compilation of high level interpreted code to optimized parallel native code.
- Develop novel HPC infrastructure in the context of combinatorics.

A key aspect will be to foster further sharing expertise and best practices between computational components.

### Description of work:

As in all other areas of science, properly supporting massively parallel architecture is a major challenge. Many of the computational components in **MathVRE** have already gone a long way in this direction. For example, an adaptation of the GAP kernel for HPC was developed during the 2009-2013 EPSRC project. The expertise gained there was then transferred to the ongoing Singular-HPC project, in particular through the rehiring of one of the developers of GAP-HPC.

In this work package, we will build on this momentum to further implement HPC support in the components Tasks 5.11, 5.12, and 5.13.

♠**TO DO:** *transition*♠

Many of the computational components of **MathVRE** use a high level interpreted language for their library. This is notably the case of Sage. Performance is achieved by compiling critical sections using the Cython Python-to-C compiler, to the expense of a lower level implementation. In Tasks 5.15 and 5.16, we will also boost performance by further developing and applying such compilation tools, while keeping a high-level approach.

### Task 5.11 *Pari*

♠**KB** [**WRITE HERE:** *Task around HPC/parallelism in Pari?*]♠

♠**TO DO:** deliverable♠

### Task 5.12 *Linbox*

♠**JGD/CP** [**WRITE HERE:** *Task around HPC/parallelism in Linbox*]♠

♠**TO DO:** deliverable♠

### Task 5.13 *Singular*

♠**WD** [**WRITE HERE:** *Task around HPC/parallelism in Singular*]♠

♠**TO DO:** deliverable♠

### Task 5.14 *HPC infrastructure for combinatorics*

♠**FH** [**WRITE HERE:** *Task around HPC infrastructure for combinatorics*]♠ ♠**TO DO:** deliverable♠

### **Task 5.15 Pythran-Cython convergence**

*Pythran is a Python to C++ compiler for a subset of the Python language. It is meant to efficiently compile scientific programs, and takes advantage of multi-cores and SIMD instruction units. Thanks to type inference, it requires little annotations. Its runtime supports a subset of the Numpy package.*

*Cython is a Python to C compiler that was originally developed for Sage and is now a thriving project of its own. It can handle essentially any Python code, and in particular classes, but relies heavily on annotations for producing optimized code.*

*Therefore, Pythran and Cython are similar in spirit but have complementary feature sets: Pythran can heavily optimize high level Numpy constructs and Cython has broader Python support. In this task, we will investigate the opportunity and feasibility of a convergence between Cython and Pythran: depending on the code at hand, one strategy or the other would be automatically selected, eventually using Pythran generated code called from Cython when relevant D 5.8. This would result in compiler-runtime cooperation driven by the Cython compiler thanks to part of the Pythran-runtime and the extra typing information provided by Cython. An effort will be made to improve more and more the parallelism in the Pythran runtime D 5.9.*

*This work will be achieved through a close collaboration between the Pythran developers hired for MathVRE and Cython developers involved in the Sage project. It should quicken Sage execution time at least on Numpy centric codes, while not putting an extra burden on the developers. Preliminary discussions with the Cython community have already taken place and received a very favorable feedback.*

### **Task 5.16 Pythran for Sage and Sage Users**

*Currently, Sage doesn't provide facilities to improve user written Python code without the modifications implied by the use of the Cython compiler. As Pythran doesn't need these codes to be rewritten, a notebook interface to compile Pythran compliant code will be added in Sage to improve user kernels using the Pythran compiler D 5.10.*

*In a similar perspective, testing and improving the integration between mpi4py and Pythran could provide an efficient toolchain for HPC while keeping full backward compatibility with pure Python code. This will require a continuous integration of Pythran to ensure its capabilities D 5.11.*

*Internally, Sage uses Cython for compiling the critical sections of its libraries. In this task, we will explore opportunities to benefit from Pythran compilation within the Sage library, in particular toward better support for parallelism. A specific challenge is that the Sage library uses quite heavily object-oriented programming.*

*This task will strongly benefit from Task 5.15, while providing in return a real life large-scale use case for it.*

*A first step to support object-oriented programming will be to make Pythran type inference more accurate, which will also improve error feedback provided for the user D 5.12.*

### **Task 5.17 Explorative task: Add support for classes in Pythran**

*Classes support is a real challenge for Pythran as it requires a more accurate typing information but also invalidates some compiler optimisations.*

*As it will need a full rework of the aliases analysis in Pythran, which is the keystone of Pythran, we are not sure it could really be integrated but it would be a proof of scalability for Pythran. Thanks to this typing and this better aliasing analysis, we could add more optimizations like the ones from Cython enabled with decorator annotation.*

**Deliverables:**

**Deliverable 5.8** *(Month 6) Implement Pythran runtime support in Cython when they are implemented instead of using default implementation.*

**Deliverable 5.9** *(Month 3) Improve Pythran runtime support to automatically take advantage of multi-cores and SIMD instruction units.*

**Deliverable 5.10** *(Month 2) Facility to compile Pythran compliant user kernels.*

**Deliverable 5.11** *(Month 1) Ensure inteoperability of Pythran with Python and its packages.*

**Deliverable 5.12** *(Month 12) Make Pythran typing better to improve error information.*



## Work package description (WP6)

Work package number	WP6	Start date or starting event:				Month 1	
Work package title	Next generation mathematical Data/Knowledge/Software-bases						
Participant number	13						
Participant short name	UZH						
Person-months per participant:	1						

**Objectives:** The objectives of WP6 is to design interfaces that can be used for a wide range of mathematical data, facilities for accepting contributions while tracking provenance and credit, and definition of standard metadata allowing for interoperability, database discovery, versioning allowing stable references, and easy recomputation.

### Description of work:

Mathematics is the only science that has not yet benefited massively from the systematic interchange of data. At the same time, mathematics has a richer notion of data than other disciplines. Indeed, "mathematical data" consists of three kinds of objects:

[D]: proper (numeric/symbolic) data

[K]: the knowledge about the mathematical objects given as statements (definitions, theorems or proofs; either formal or rigorously informal)

[S] : software that computes (with) the mathematical objects

All three kinds of "data" are equally important for mathematics and are tightly interlinked:

[D] serves as examples for [K] or as counterexamples for conjectures in [K];

[S] computes [D] and establishes properties of [D] (given as [K]);

[D] tests [S], [S] is verified wrt. [K];

theorems and proofs in [K] induce and justify algorithms for [S];

[D] induces conjectures and guides proofs in [K].

Many mathematical databases now exist, but their internal structure does not surface this richness. This prevents the formulation of new conjectures, the testing of new hypotheses, and generally an exploratory approach to mathematical data. The past has shown that this approach could be fruitful:

- both the Riemann Hypothesis and the Birch and Swinnerton Dyer conjectures resulted from exploratory  $L$ -function computations, and now stand among the seven Clay Millennium Problems;
- the Monstrous Moonshine conjecture finds its origin in a numerical coincidence between dimensions of representations of the Monster group and coefficients of the  $j$ -function, and its conclusion eventually led to Borcherds' Fields medal.

Therefore we need ways to represent DKS in the same systems, and – since DKS get huge in current computational/experimental mathematics – we need a new kind of "databases", which we will call MDKS-bases (Mathematical Data/Knowledge/Software-bases).

This complexity is on vivid display in the *L-functions and Modular Forms database* project (LMFDB): while the general shape of the functional equation of an  $L$ -function is dependent on a lot of theoretical knowledge, it also requires parameter data and the coefficients of the associated Dirichlet series. Once this is obtained, highly optimised (and heavily parallelizable) algorithms can be run to compute values of this function.

We propose in this work package to design and build an infrastructure that would make it easy for either isolated mathematicians or a distributed collaboration to manage mathematical data. This work would provide part of the backend to Work Packages **♠TO DO: work package on interfaces, and???**♠, and would draw on previous work with the LMFDB and FindStat (which will be treated as prototypes for our purposes). Requirements should be as minimal as possible (depending on the individual's goals), and in particular would not require any background in databases to contribute new data or perform queries.

### Task 6.18 Survey of existing databases

*All the systems considered in this proposal (GAP, Sage, Pari, Singular) include data as part of their regular distribution. In this task, we will survey existing databases, the technology used to implement them, how they were linked to the rest of the existing infrastructure and the functionalities offered. We will also select additional external data and projects to add to this effort, aiming to maximise the impact of our work.*

### Task 6.19 Design of new infrastructure, formulation of requirements

*Ontologies are the canonical method used to implement databases that require significant data interchange. However,*

because of extreme reification in mathematics, this is not entirely suitable for our goals. We will design a new infrastructure for **MathVRE**, drawing on existing emerging standards .

#### **Task 6.20 Triform Theories in OMDoc/MMT**

OMDoc/MMT is a representation language for mathematical knowledge and documents. Carette and Farmer have developed the notion of biform theories (K/S) in a uniform representational approach; our work here would extend this along the data axis, which will require a specialised but integrated treatment.

#### **Task 6.21 Computational Foundation for Python/Sage (or some CAS)**

In the OMDoc/MMT world a foundation is a logical base language that gives the formal meaning to all objects represented/formalized in it. We have created a very initial computational foundation for Scala and implemented it in the MMT API. This can be used to execute (or verify) computations directly in OMDoc/MMT and thus forms the basis for various integration tasks for OMDoc/MMT biform theories that integrate Scala computations. Here we propose to develop a somewhat more complete computational foundation for Python and/or parts of Sage (coverage to be determined). Bi/Triform theories come in three parts:

- syntax: what operators/types are there, how do they nest,
- computation: what does the computation relation look like (sometimes called operational semantics). The declarative semantics of a computational foundation can be given as an OMDoc/MMT theory morphism into another foundation (e.g. a set theory);
- ??? three parts

#### **Task 6.22 LMFDB Case study (Triformal Theories)**

In this task we would develop triformal theories for an exemplary part of the LFMDB to test the design from K6.5. We will identify a fragment of the LFMDB that we want to model and design the model. Then we will perform cross-validation of the three model parts against each other (essentially model-based testing of software and inference). Finally, we will pick an algorithm from the LFMDB and verify it against its specification and the computational foundation developed in K6.4. (decrease importance of verification as opposed to interoperability)

This Task will be co-developed with K6.4, it will validate the design of triformal theories and be iterated to test the design changes.

#### **Task 6.23 OEIS Case Study (Coverage and automated Import)**

In this case study we test the practical coverage of the trifunctional modules, by transforming an existing, high-profile database (the Online Sequence of Integer Sequences <http://www.oeis.org>) into OMDoc/MMT. The OEIS has about 250 thousand sequences, with formulae, descriptions, definitions, references, software, etc. in a structured text file (but no standardized format for formulae and references), so we expect to get 250 k theories. Having the OEIS in OMDoc/MMT form allows to do Knowledge Management services (presentation, definition lookup, formula search, ...) in MathHub.info (see WP4.?). The OEIS is a good case study, since the DKM are licensed under a CC license which allows derived works. The large size will allow statistically significant semantic cross-validation of the heuristic transformation process and thus achieve a significant DKS community resource.

#### **Task 6.24 Findstat Case Study ???**

#### **Task 6.25 Memoization**

## Deliverables:

- Conversion of existing and new databases to unified interoperable system:
  - Polytopes in Polymake
  - graphs, graph properties
  - LMFDB data
  - FindStat data
  - Finite groups (Max)
  - Lattices
- ?? (Month X): Shared persistent memoization library for Python/Sage. Typical use case: A group of collaborators is using intensively a given function (in Sage, or in their private code). They want to memoize the results, as with e.g. Sage's `cached_method`, but across sessions. They further want to share the underlying growing database between themselves, and maybe eventually publish it.

### Features:

- Use, further extend, and contribute back to some established (Python?) persistent memoization infrastructure. E.g.
  - \* <https://pythonhosted.org/joblib/memory.html>
  - \* [github.com/vivekn/redis-simple-cache](https://github.com/vivekn/redis-simple-cache)
  - \* [bitbucket.org/zzzeek/dogpile.cache](https://bitbucket.org/zzzeek/dogpile.cache)
- Apply not only to user-level functions, but also to lower level functions, e.g. in the Sage library, so that indirect calls to the function also get memoized.
- Trivial to setup and configure for the end user: in a single line, the user selects an existing function, a backend (with a default value), maybe provide some semantic information, and voila.  
Typical interface: a decorator to be set on appropriate functions. ♦**TO DO: Mock code**♦
- Trivial to setup and configure for groups of researchers, with a wide range of storage backends (e.g. shared dropbox folder, remote directory, database, git repository, ...).
- Easy to setup data-bot: e.g. launching a virtual machine that systematically fills up the shared database.
- Versioning and provenance tracking (user, algorithm, software version, ...), for quality certification, credit, ...
- Recomputation?
- Ease of publishing, importing, ...
- Usual database properties: atomicity, merging (easy since the results are supposed to be immutable: just need to merge the tracking info), alerts in case of divergence.

- D6.2: workshop report
- D6.3.1: Design of Triform (DKS) Theories (Specification/RNC Schema/Examples) 12 PM
- D6.3.2: Implementation of Triform Theories in the MMT API.
- D6.4.1: Python/Sage Syntax Foundation Module in OMDoc/MMT 12PM
- D6.4.2: Python/Sage Computational Foundation Module in OMDoc/MMT
- D6.4.3: Python/Sage Declarative Semantics in OMDoc/MMT
- D6.5.1: LFMDB deep modeling: Fragment Identification Initial Model Design 12PM
- D6.5.2: LFMDB Data vs. Knowledge vs. Software Validation
- D6.5.3: LFMDB Algorithm verification wrt. a Triformal theory

MathVR E.6.5.4: LMFDB full integration of algorithms, data and presentation (not so much verification) December 15, 2014

- D6.6.1: Heuristical Parser for the OEIS

D6.6.2: Group Validation for OEIS DKS Theories

=====8<-----

Another connection: on several occasions, we found that software was the best way to represent certain databases of mathematical knowledge. E.g. in Algebraic Combinatorics we have a whole zoo of Hopf algebras. Many of them are implemented in MuPAD/Sage by specifying the objects that index the basis together with computation rules for the product and coproduct. When we want to retrieve information about such algebras, it's usually much more convenient to look at the code than to search through the literature. Especially since the code is usually more correct than the literature because it's *\*tested\**.

We may also think of providing an interface to LMFDB via SCSCP protocol (<http://www.symbolic-computing.org/scscp>) so it may be accessed by a variety of other systems (see their current list at <http://www.symbolic-computing.org/scscp>)

database access to LMFDB as a python library

## Work package description (WP7)

<b>Work package number</b>	WP7	<b>Start date or starting event:</b>	Month 1
<b>Work package title</b>	<b>Development Models for an Academic Free Software Ecosystem</b>		
<b>Participant number</b>	9		
<b>Participant short name</b>	USTAN		
<b>Person-months per participant:</b>	1		

**Objectives:** The objectives of WP7 are to:

- 
- 
- 
- 
- 

**Description of work:**

This workpackage ...

**Deliverables:**

- ?? (Month X): X.

♠DP/UM [WRITE HERE: Workpackage Social Aspects]♠

## Work package description (WP8)

Work package number	WP8	Start date or starting event:					Month 1
Work package title	Social Aspects						
Participant number	6						
Participant short name	UO						
Person-months per participant:	1						

**Objectives:** The objectives of WP8 are to:

- 
- 
- 
- 
- Supporting the Mathematical Process

**Description of work:**

This workpackage ...

**Deliverables:**

**Deliverable 8.13** (*Month 12*) ...

## Work package description (WP9)

<b>Work package number</b>	WP9	<b>Start date or starting event:</b>	Month 1
<b>Work package title</b>	<b>Dissemination, Exploitation and Communication</b>		
<b>Participant number</b>	9		
<b>Participant short name</b>	USTAN		
<b>Person-months per participant:</b>	1		

**Objectives:** The objective of this work package is to organize and optimize the communication with the larger community. This includes:

- reviewing emerging technologies;
- disseminating research results to the scientific community;
- ensuring awareness of the results in the user community;
- raising general public awareness of the **MathVRE** project;
- defining individual exploitation plans; and,
- managing existing and new intellectual property.

### **Description of work:**

Dissemination: software, APIs, technologies, research results, ...

### **Task 9.26 Reviewing emerging technologies**

*In this task, we will produce periodic reviews of emerging technologies and relevant developments elsewhere, and implications for our plans. This include the review of standard components and service for storage and sharing, computational resources, authentication, package management, etc. This may further include negotiating access or shared development when appropriate. This information will be fed to the other work packages, in particular Work Package WP3component Architecture.*

### **Task 9.27 Dissemination and Communication activities**

♠**TO DO:** scale this down as appropriate♠

*This task comprises all forms of direct dissemination and public communication activities such as press releases, creation of the project web-site including visitor analysis and monitoring tools (9.15), scientific and technical publications, outreach activities (seminars, keynote talks, media interviews, press releases), pro- motion through social media (e.g. twitter, facebook, linkedin), technical workshop organisation, creation of advertisement materials such as flyers, posters, and electronic feeds as well as their distribution.*

*At least two press releases will be generated in the course of the project (9.14, ??), and the project will organise at least one open technical workshop each year.*



**Deliverables:****Deliverable 9.14** (Month 3) Press release announcing start of *MathVRE*.**Deliverable 9.15** (Month 3) Project web site**Deliverable 9.16** (Month 12) Year 1 report**Deliverable 9.17** (Month 24) Year 2 report**Deliverable 9.18** (Month 36) Year 3 report**Deliverable 9.19** (Month 48) Year 4 report

Raw material:

- Documentation improvements: overview, cross links, overview of recent improvements
- Thematic tutorials
- Collections of pedagogical documents  
E.g. a complete collection of interactive class notes with computer lab projects for the “Algèbre et Calcul formel” option of the French math aggregation (starting from 2014-2015, only open-source systems will be supported, and Sage is a major player).
- Localization of the Sage user interface and key documents in various European languages.
- Distribution of the documents either in the main distribution of Sage or through the online repository (see collaborative tools).
- Massive online introduction course to Sage, drawing on the sage tutorial/notebooks. Could be “First year Sage course in a box”.
- Taking the opportunity of Python courses to propose Sage as a natural extension for mathematics; an example is French’s “Classes préparatoires”<sup>1</sup>, where Python has been recently selected as the language to learn programming<sup>2</sup>.

♠**TO DO:** *Milestones need to be discussed and then described here.*♠

<sup>1</sup>[http://en.wikipedia.org/wiki/Classe\\_pr\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{e\global\mathchardef\accent@spacefactor\spacefactor}\accent19\egroup\spacefactor\accent@spacefactorparatoire\\_aux\\_grandes\\_\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{e\global\mathchardef\accent@spacefactor\spacefactor}\accent19\egroup\spacefactor\accent@spacefactorcoles](http://en.wikipedia.org/wiki/Classe_pr\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{e\global\mathchardef\accent@spacefactor\spacefactor}\accent19\egroup\spacefactor\accent@spacefactorparatoire_aux_grandes_\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{e\global\mathchardef\accent@spacefactor\spacefactor}\accent19\egroup\spacefactor\accent@spacefactorcoles)

<sup>2</sup>See the “Annexe” at [http://www.education.gouv.fr/pid25535/bulletin\\_officiel.html?cid\\_bo=71586](http://www.education.gouv.fr/pid25535/bulletin_officiel.html?cid_bo=71586)

♠**TO DO:** *Check this for any necessary changes.*♠

## 3.2 Management Structure and Procedures

♠**EC Commentary:** *Will get help from Orsay's grant services*♠

♠**EC Commentary:** *Describe the organisational structure and the decision-making (including a list of milestones (table 3.2a)).*

*Explain why the organisational structure and decision-making mechanisms are appropriate to the complexity and scale of the project.*

*Describe, where relevant, how effective innovation management will be addressed in the management structure and work plan.*

*Describe any critical risks, relating to project implementation, that the stated project's objectives may not be achieved. Detail any risk mitigation measures. Please provide a table with critical risks identified and mitigating actions (table 3.2b).*♠

### 3.3 Consortium as a Whole

#### ♠EC Commentary:

- *Describe the consortium. How will it match the project's objectives? How do the members complement one another (and cover the value chain, where appropriate)? In what way does each of them contribute to the project? How will they be able to work effectively together?*
- *If applicable, describe the industrial/commercial involvement in the project to ensure exploitation of the results and explain why this is consistent with and will help to achieve the specific measures which are proposed for exploitation of the results of the project (see section 2.3).*
- *Other countries: If one or more of the participants requesting EU funding is based in a country that is not automatically eligible for such funding (entities from Member States of the EU, from Associated Countries and from one of the countries in the exhaustive list included in General Annex A of the work programme are automatically eligible for EU funding), explain why the participation of the entity in question is essential to carrying out the project*

♠

♠**TO DO:** *The participants are core developers of the involved components*♠

♠**TO DO:** *Experience in community building and engagement*♠

Writing interfaces between computer algebra systems from different areas and collaborative software development are important themes within the DFG Priority Project SPP1489. As in the SAGE community, networking measures include the regular exchange of developers and the regular organization of software workshops (coding sprints) which bring whole teams together for solution finding and intense code writing. Particular tight collaborations exist between the GAP and the SINGULAR communities, with major GAP-SINGULAR developers meetings taking alternately place at St. Andrews, Kaiserslautern, and Aachen. See <http://www.computeralgebra.de/>.

♠**TO DO:** *User interfaces: recruitment of IPython developers*♠

♠**TO DO:** *Pythran and HPC: a key asset will be the recruitment of two of the lead developers of the Pythran Python-to-C compiler.*♠

♠JGD/CP [**WRITE HERE:** Linbox: recruitment of previous ANR developer]♠

♠**TO DO:** *Explanation of why we want to include Seattle (sage-math cloud, is a key component; access to IP).*♠

### 3.4 Resources to be Committed

♠**EC Commentary:** *Will get help from Orsay's grant services♠*

♠**EC Commentary:** *Please provide the following:*

- *a table showing number of person/months required (table 3.4a)*
- *a table showing 'other direct costs' (table 3.4b) for participants where those costs exceed 15% of the personnel costs (according to the budget table in section 3 of the administrative proposal forms)*

♠

#### Summary of staff effort

♠**EC Commentary:** *Please indicate the number of person/months over the whole duration of the planned work, for each work package, for each participant. Identify the work-package leader for each WP by showing the relevant person-month figure in bold.♠*

♠**TO DO:** *Update this once the list of parthers and the WPs are finalised.♠*

Partic. no.	Partic. short name	Work package									Total PMs
		WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	
1	UPS										
2	Logilab										
3	UVSQ										
4	UJF										
5	UB										
6	UO										
7	USHEF										
Total PM											

♠**EC Commentary:** *Please complete the table below for each participant if the sum of the costs for 'travel', 'equipment', and 'goods and services' exceeds 15% of the budget table in section 3 of the proposal administrative forms).♠*

#### Other direct cost items

	Cost (€)	Justification
Travel		
Equipment		
Other goods and services		
Total		

#### Management Level Description of Resources and Budget

♠**TO DO:** *This needs to be updated in line with the rest of the project.♠*

The project will employ XX person-months of effort over YY years, comprising ...

♠**EC Commentary:** *This section is not covered by the page limit.  
The information provided here will be used to judge the operational capacity.*♠

## 4 Members of the Consortium

### 4.1 Participants

♠**EC Commentary:** *Please provide, for each participant, the following (if available):*

- *a description of the legal entity and its main tasks, with an explanation of how its profile matches the tasks in the proposal;*
- *a curriculum vitae or description of the profile of the persons, including their gender, who will be primarily responsible for carrying out the proposed research and/or innovation activities; this includes a description of the profile of the to-be-recruited personnel*
- *a list of up to 5 relevant publications, and/or products, services (including widely-used datasets or software), or other achievements relevant to the call content;*
- *a list of up to 5 relevant previous projects or activities, connected to the subject of this proposal;*
- *a description of any significant infrastructure and/or any major items of technical equipment, relevant to the proposed work;*
- *any other supporting documents specified in the work programme for this call.*

♠

♠**SL [WRITE HERE: Saint Andrews]**♠

## Université Paris Saclay

♠**TO DO:** *INRIA or INRA below?* – sorry to jump in, Paul♠

University Paris-Sud is among the 40 top universities worldwide in the 2013 Shanghai ranking, and is one of the two best French research universities. With about 27000 students, 1800 permanent teaching staff and 1300 permanent research scientists from national research organisations (CNRS, Inserm, INRA), it is the largest campus in France. Since 2006, scientists from the University were awarded two Fields medals, one Nobel Prize and a number of other international (European Inventor Award 2013, Wolf Prize 2010, Holweck Prize 2009, Japan prize 2007) and national prizes. The Université Paris-Sud has a complete array of competences, ranging from the purest of exact sciences to clinical practices in medicine, covering life and health sciences, legal sciences and economics. Research at the Université Paris-Sud, an essential part of academic understanding, is complemented by research activities with a high valorisation potential. Research contracts and partnership with companies make the Université Paris-Sud a key actor and a major player in French research. The University is located close to the Plateau de Saclay, the largest cluster of public and private R&D institutions in France (with ca. 16000 research staff), and is one of the core members of the University Paris Saclay – a world class university and a world-renowned research and innovation hub.

In the context of this project, the Université Paris Saclay is the home of one of the largest group of Sage developers worldwide. It's a member of the Open Source Thematic Group of the Systematic Paris Region Systems and ICT Cluster. The University also hosts a major research group working on proof assistants (Coq), which naturally opens the door for reaching toward this neighbor community.

### Curriculum vitae of the investigators

**Florent Hivert** Professor at the Laboratoire de Recherche en Informatique, Florent Hivert is a senior researcher in Algebraic Combinatorics with 29 papers in international journals and 15 communications in international conferences. With 100 tickets (co)authored and as many refereed, Hivert is himself a core Sage developer, with contributions including key components of the Sage infrastructure (documentation, automated test, combinatorics infrastructure, parallelism, ...), specialized research libraries.

**Viviane Pons** Maître de Conférences at the Laboratoire de Recherche en Informatique, Viviane Pons is a young researcher in Algebraic Combinatorics. She defended her thesis in 2013 and has 3 papers in international journals and 3 communications in international conferences. Before committing herself to research, she spent two years working in industry as a Java and web developer.

She discovered Sage during her first sage-days in 2010 and has since been an active user and contributor with 10 (co)authored tickets improving the support of combinatorial objects into Sage. She is very involved in the promotion of Sage, participating in Sage-Days and proposing Sage introduction tutorials or Sage presentations in various conferences. She is also one of the main developers of the project FindStat dedicated to databases in combinatorics.

**Nicolas M. Thiéry** Professor at the Laboratoire de Recherche en Informatique, Nicolas M. Thiéry is a senior researcher in Algebraic Combinatorics with 15 papers published in international journals. Among other things, he is a member of the permanent committee of FPSAC, the main international conference of the domain, and has collaborators in Canada, India, and in the US where he spent three years (Colorado School of Mines, UC Davis); he also coorganized fourteen international workshops, in particular Sage Days, and the semester long program on "Automorphic Forms, Combinatorial Representation Theory and Multiple Dirichlet Series" hosted in Providence (RI, USA) by the Institute for Computational and Experimental Research in Mathematics.

Algebraic combinatorics is a field at the frontier between mathematics and computer science, with heavy needs for computer exploration. Pioneer in community-developed open source software for research in this field, Thiéry founded in 2000 the Sage-Combinat software project; with 50 researchers in Europe and abroad, this project has grown under his leadership to be one of the largest organized community of Sage developers, gaining a leading position in its field, and making a major impact on one hundred publications<sup>3</sup>. Along the way, he coauthored part of the proposal for NSF Sage-Combinat grant OCI-1147247.

<sup>3</sup><http://sagemath.org/library-publications-combinat.html>, <http://sagemath.org/library-publications-mupad.html>

With 150 tickets (co)authored and as many refereed, Thiéry is himself a core Sage developer, with contributions including key components of the Sage infrastructure (e.g. categories), specialized research libraries (e.g. root systems), thematic tutorials, and two chapters of the book “Calcul Mathématique avec Sage”.

### Publications, achievements

♠**TO DO:** *Il faut être plus formel dans la description des projets antérieurs : Acronyme, titre, agence de financement, durée. Pareil pour les publi - auteurs, titre exact, année etc.*♠

1. Lead of the Sage-Combinat software project.
2. Coauthoring of the open source book “Calcul Mathématique avec Sage”, the first of its kind comprehensive introduction to computational mathematics in Sage for education.
3. XXX tickets contributed to Sage.

### Previous projects or activities

1. Home of six one week-long Sage Days workshops.
2. Co-Organizer of ♠**TO DO:** XXX♠ Sage Days.
3. Founder and regular organizer of a bimonthly Sage User Group meeting in the greater Paris area.
4. Expertise exchanges with Logilab
5. ♠**TO DO:** XXX♠

### Significant infrastructure

The Université Paris Sud hosts the lead developers of the open source cloud infrastructure Stratuslab and its reference infrastructure (♠**TO DO:** XXX cores♠). The participants are regular users of this infrastructure, and in close contact with the developers.

♠**TO DO:** *Comments by Olivier Chapuis*♠ Paris Sud also hosts the WILDER platform, an experimental wall-sized high-resolution interactive touch-screen for conducting research on collaborative human-computer interaction and the visualization of large datasets.

#### 4.1.1 University of Kaiserslautern

**Principal investigator Prof. Dr. Wolfram Decker** Wolfram Decker is a professor of mathematics at TU Kaiserslautern. He formerly was a research fellow at Berkeley with a NATO grant, a visiting researcher at Kyoto with a JSPS grant, and a professor at Saarbrücken, Germany. Decker has more than thirty publications including two books on computational algebraic geometry and papers in *Compositio*, *Crelle*, and *Mathematische Annalen*. He has held several grants in four different priority programmes of the German Research Council DFG and is now coordinator of the priority programme SPP 1489 “Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory”. He was also coordinator of the European algebraic geometry network EuroProj (1996–1999) and Chair of the programme management committee of the European algebraic geometry network EAGER (2000–2004). He held seven grants for EU Highlevel Scientific Conferences and (co-)organized about 50 conferences, summer schools, workshops, and coding sprints. He was Chair of the Minisymposium on Computer Algebra during the third ECM. Decker has supervised 13 PhD students. He has been a frequent lecturer at the African Institute of Mathematics (AIMS) at Cape Town, and he has run 8 schools on computational algebraic geometry in different countries.

Decker’s research interests lie in areas of algebraic geometry and computer algebra. In addition to writing theoretical papers, he is a leader in mathematical software development and has written thousands of lines of code himself. He has made contributions to the systems MACAULAY2 and, much more substantially, SINGULAR. Since 2009 he is the head of the SINGULAR development team. Current tasks of the team include crosslinking SINGULAR to other systems, most notably to GAP, and parallelizing SINGULAR. These tasks are fundamental to the **MathVRE** project.

#### Publications, products, achievements

1. Head of SINGULAR development team.
2. Coordinator of the DFG Priority Project SPP1489 *Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory*.
3. Coauthor of the book *Computing in Algebraic Geometry - A Quick Start using Singular*, Springer.
4. Coauthor of the book *A First Course in Computational Algebraic Geometry*, Cambridge University Press.

#### Previous projects or activities

1. Member of the DFG Priority Project *Algorithmic Number Theory and Algebra*.

**Significant infrastructure** Excellent computing infrastructure (high end servers), access to different types of compute clusters through the IT-Center of the TU Kaiserslautern.



## University of Oxford

♠DP/UM [WRITE HERE: Description of the university of Oxford]♠

## Curriculum vitae

♠UM [WRITE HERE: CV Ursula Martin]♠ ♠DP [WRITE HERE: CV Dima Pasechnik]♠

## Publications, products, achievements

1. ♠DP/UM [WRITE HERE: Publications Oxford]♠

## Previous projects or activities

1. ♠DP/UM [WRITE HERE: Projects and activities in Oxford]♠

## Significant infrastructure

♠DP/UM [WRITE HERE: Significant infrastructure in Oxford]♠

## Université de Bordeaux

- INRIA, LaBRI, IMB
- journal de théorie des nombres
- Plafirm and a mesocentre Avakas
- Several softwares developed in Bordeaux: pari/GP, tulip, etc

## Curriculum vitae

**Vincent Delecroix** CNRS researcher at the Laboratoire Bordelais de recherche en informatique, Vincent Delecroix is a junior researcher in Number theory, Combinatorics and Dynamical systems.

X publications. He has several international collaboration (England, Mexico, United-States).

- Bobo - cours Sage Bordeaux

Delecroix is a major developer of Sage in various components: - integers - combinatorics - dynamical systems

**Karim Belabas** Karim is One of the main pari developer.

**Bill Allombert** CNRS Ingénieur de Recherche. One of the main pari developer.

## Publications, products, achievements

Some recent Publications :

1. Belabas, Karim; Friedman, Eduardo; Computing the residue of the Dedekind zeta function. Math. Comp. 84 (2015), no. 291, 357–369.
2. The PARI Group; PARI/GP version 2.7.0, Bordeaux, 2014, <http://pari.math.u-bordeaux.fr/>.
3. Belabas, Karim et al. Explicit methods in number theory. Rational points and Diophantine equations, 179 pages, Panoramas et Synthèses 36, 179p., 2012.
4. Allombert, Bill; Bilu, Yuri and Pizarro-Madariaga, Amalia;) CM-Points on Straight Lines , to appear in "Analytic Number Theory" (dedicated do H. Maier), Springer.

## Previous projects or activities

Current grants:

1. ANR PEACE (2012-2015) Goal: The discrete logarithm problem on algebraic curves is one of the rare contact points between deep theoretical questions in arithmetic geometry and every day applications. On the one side it involves a better understanding, from an effective point of view, of moduli space of curves, of abelian varieties, the maps that link these spaces and the objects they classify. On the other side, new and efficient algorithms to compute the discrete logarithm problem would have dramatic consequences on the security and efficiency of already deployed cryptographic devices.
2. ERC starting grant ANTICS (2011-2016) Goal: "Rebuild algorithmic number theory on the firm grounds of theoretical computer science". Challenges: complexity (how fast can an algorithm be?), reliability (how correct should an algorithm be?), parallelisation.

## Significant infrastructure

♠VD [WRITE HERE: this still needs to be done]♠

Two center for computations: Plafirm and Avakas.

### Curriculum vitae

**Jean-Guillaume Dumas** Professor at the Laboratoire Jean Kuntzmann, Jean-Guillaume Dumas is a senior researcher in Computer Algebra with 40 papers published in international journals or refereed international conferences. Among other things, he is vice-president of ACM Special interest group on symbolic and algebraic manipulations (SIGSAM), department chair within his Laboratoire (6 research teams, 130 members) and has collaborators in USA, Canada, Ireland, Germany and Luxembourg; he has also co-organized fifteen international conferences.

Computer Algebra is a field at the frontier between mathematics and computer science, with heavy needs for computer exploration. Jean-Guillaume Dumas is the main developer of the LinBox and Givaro C++ libraries (libgivaro1, libgivaro-dev, libgivaro-doc, liblinbox0, liblinbox-dev in Debian) used, e.g., by Sage respectively as its exact linear algebra and its finite fields.

Along the way, he coauthored part of the proposal for NSF-INRIA grant QOLAPS on Quantifier elimination, Optimization, Linear Algebra and Polynomial Systems and he is the director of the French ANR program on High-Performance Algebraic Computations.

**Clément Pernet** Associate Professor at the joint Inria-LIG research group MOAIS, Clément Pernet is a junior researcher in Computer Algebra, parallel computing and coding theory with 16 papers published in international journals or refereed international conferences. He is associate editor of the ACM transactions on Mathematical Software and has co-organized 10 conferences, including 2 sage-days and the 2012 edition of ISSAC, the leading conference in computer algebra.

Since he was a post-doc at University of Washington, under the supervision of William Stein, head of the Sage project, he has had many contributions to Sage on the exact linear algebra and the symbolic computation tools. He co-authored the book “Calcul Mathématique avec Sage” with the chapter on Linear algebra. Clément Pernet is the founder and lead developer of the fflas-ffpack library, kernel for dense linear algebra over a finite field, delivering high performance computation to LinBox and Sage. He is a core contributor to the LinBox library and contributed to the m4ri library.

**Pierrick Brunet** Junior Research and Development Engineer at INRIA Grenoble, Pierrick Brunet is working on compilation of C/C++ OpenMP program to C/C++ programs with calls to specific OpenMP runtimes.

With about 25% of commits in the Pythran project, Pierrick is one of the core devs of this project which compile a subset of the Python language to native Python modules.

### Publications, products, achievements

#### Software projects

**fflas-ffpack:** An open-source C++ library offering dense linear algebra kernels over a finite field. In the same spirit as the numerical BLAS (Basic Linear Algebra Subroutines), and LAPACK libraries, it delivers high performance for the most commonly used routines of scientific computing: matrix multiplication, solving linear systems, computing echelon forms, determinants, characteristic polynomials, etc. This library has set the standard approach for high performance exact dense linear algebra. It is currently used in Sage, and has inspired the design of similar routines in most commercial computer algebra softwares: maple, magma, etc.

**LinBox:** An open-source C++ middleware library for exact linear algebra. It uses fflas-ffpack for its dense finite field linear algebra part and extends its functionalities to other computation domains (integers, rationals, polynomial rings) and type matrices (sparse and structures matrices, black-box matrices). LinBox is integrated in Sage.

**Pythran:** An open-source Python-to-C++ optimizing compiler offering an high performance runtime for Scientific Python kernels. Dynamicity of the Python language is not compliant with static compilation. That’s why only a subset of the Python language is supported by Pythran. Thanks to these restrictions, Pythran generate code up to 3000 faster than original module.

#### Selected Publications

1. Coauthoring of the open source book “Calcul Mathématique avec Sage”, the first of its kind comprehensive introduction to computational mathematics in Sage for education.
2. Parallel computation of echelon forms (with J-G. Dumas, T. Gautier and Z. Sultan). *In Proc. Euro-Par’14* (2014), LNCS 499–510. DOI: 10.1007/978-3-319-09873-9\_42.
3. Pythran: Enabling static optimization of scientific python programs (Serge Guelton, Pierrick Brunet, Alan Raynaud, Adrien Merlini, and Mehdi Amini.) *Proceedings of the Python for Scientific Computing Conference (SciPy)* June 2013.
4. Fast Computation of Hermite Normal forms of random integer matrices (with W. Stein). *J. of Number Theory* **130.7** (2010), 1675–16833. DOI: 10.1016/j.jnt.2010.01.017
5. Dense Linear Algebra over Word-size Prime Fields (with J.-G. Dumas and P. Giorgi). *Trans. on Math. Software* **35.3** (2008), 1–42. DOI: 10.1145/1391989.1391992.
6. Faster Computation of the Characteristic Polynomial (with A. Storjohann). *In Proc. ISSAC’07* (2007), 307–314. DOI: 10.1145/1277548.

### Previous projects or activities

1. Direction of the ANR program on High-Performance Algebraic Computations 2012-2015.
2. Participation to the NSF-Inria associate teams QOLAPS (with NCSU, USA)
3. Coordination of a CNRS PEPS grant (parallel computer algebra)
4. Organization of the ISSAC’12 conference, the main international conference in computer algebra, and of PASCO’15 a satellite conference on parallel computer algebra.

### Significant infrastructure

♠JGD/CP [WRITE HERE: Significant infrastructure in Grenoble (or remove section)]♠

## Logilab

Logilab (<http://www.logilab.fr/>) is a french SME focused on using the web and free software to help scientists. It has been in business since 2000 and counts over 20 engineers and PhDs proficient in software engineering, knowledge representation, design and management of IT infrastructure, etc.

Logilab invests 15% of its turnover in research and development and has been part of several R&D projects at the national and european levels, always to provide technical expertise and support to the other partners.

In the context of this project, Logilab will innovate to support the partners with tools and infrastructure, including open databases to flexibly store mathematical objects, user interfaces to visualize complex mathematical properties, fluid workflow tools to ease large-scale collaboration, etc.

Logilab's PIC number is 948455525.

## Curriculum vitae

### Publications, products, achievements

1. ♠XXX [WRITE HERE: ...]♠
2. CubicWeb (mention of prize at DataConnexion#3)
3. publi Brainomics

### Previous projects or activities

1. ♠XXX [WRITE HERE: ...]♠
2. ASWAD (eu)
3. KIDDANET (eu)
4. PYPY (eu)
5. OpenHPC (fr/FUI)
6. BRAINOMICS (fr/ANR)

### Significant infrastructure

♠XXX [WRITE HERE: ...]♠

## University of Warwick

The Mathematics Institute at the University of Warwick was ranked 23rd worldwide in the 2013 QS world university subject rankings. Five members of the Department are Fellows of the Royal Society, and one, Regius Professor Martin Hairer, was awarded a Fields Medal in 2014. Mathematics and Statistics at Warwick currently hold £35.8M in research grants from EPSRC (the next highest in the UK being Cambridge at £22.8M and Oxford at £24.2M). Nine members of the department currently hold ERC grants.

### Curriculum vitae

**John E. Cremona** Professor of Mathematics. DPhil (Oxford, 1981) under Birch. Previous posts: Michigan, Dartmouth (US), Exeter, and Nottingham (as chair and Head of Pure Mathematics). Cremona has around 50 publications, including a book and papers in *Compositio* and *Crelle*. He has held grants from EPSRC and other UK sources worth £2.5M as well as €2.5m from the EU for Marie-Curie Research Training Networks in 2000-2004 and 2006-2010. He was a Scientist in Charge of one of twelve teams in both of these networks, and leader of the research project “Effective Cohomology Computations” in the second. He has been on the Scientific Committee of 30 international conferences (including several Sage Days), and given many invited lecture series. He co-organised semester-long research programmes at IHP Paris (2004) and MSRI (2011). He has been an editor for five journals. He has supervised 16 PhD students, a dozen Masters students, two EU-funded postdoctoral fellows and currently has three EPSRC-funded postdoctoral research assistants. Cremona has given over 30 invited conference addresses and seminars in 9 countries in the last 10 years.

Cremona’s research includes areas of particular relevance to the current project. His methods for systematically enumerating elliptic curves, which are the subject of a book and numerous papers, have been used to compile a definitive database of elliptic curves which is very widely cited, and now forms part of the LMFDB. Cremona’s experience in managing such computations and the management, publication and electronic dissemination of the resulting large datasets set a standard which large-scale number-theoretical database projects such as the LMFDB now seek to match. Cremona’s experience and reputation in this field have been important for the successful management of the LMFDB project.

Cremona has been the leading computational number theorist in the UK since his PhD thesis in 1981, following in the tradition of Birch and Swinnerton-Dyer. He has written thousands of lines of code in his C++ library *eclib* (one of the standard packages included in Sage since its inception) which includes his widely-use program *mwrnk* for computing ranks of elliptic curves. As well as writing thousands of lines of new python code for Sage, he has also contributed to the active number-theoretical packages *Pari/GP* and *Magma*.

### Publications, products, achievements

1. The Number Theory research group at Warwick was started only in 2006, but has rapidly risen to international status and one of the largest and most vibrant groups in Europe, comprising 25 members (professors, lecturers, postdoctoral researchers and early stage researchers). Of the group’s members, two (Loeffler and Dokchitser) hold Royal Society Research Fellowships and one (Bartel) a Royal Commission 1851 Fellowship. Loeffler won a Leverhulme Foundation Prize jointly with Zerbes.
2. Several members of the Number Theory group at Warwick are Sage developers, including John Cremona, who has contributed thousands of lines of code to Sage since 2006 both through his *eclib* C++ library and through original Python code which forms part of the Sage library; David Loeffler, who has contributed substantially to the modular forms module in Sage; and postdoc Marc Masdeu, who has worked on the Sage-Flint interface.

### Previous projects or activities

1. In 2013 Professors John Cremona and Samir Siksek, together with co-investigators at Bristol, were awarded a six-year major grant of £2.2M from the UK Engineering and Physical Sciences Research Council (EPSRC) to support the L-functions and Modular Forms Database (LMFDB) project. This grant funds three postdoctoral researchers at Warwick, computer equipment to host its database and website, and regular LMFDB workshops.
2. Each year Warwick hosts a year-long Warwick EPSRC Symposium focussing on one area of mathematical research. The 2012-13 Number Theory Symposium included six research workshops and a summer school

“Number Theory for Cryptography” and raised the international profile of the number theory group substantially.

### **Significant infrastructure**

Computing infrastructure available to the group is excellent, with seven dedicated machines (over 300 cores) as well as access through Warwick’s Centre for Scientific Computing which hosts a 6000-core linux cluster and a 3500-core cluster of workstations.

## University of Zürich

The University of Zurich consistently ranks among the top 15 research institutions in Europe. It is the largest university in Switzerland, with over 26000 students, and offers the most comprehensive academic program of the country. It has close to 600 professors and over 5000 academic staff.

Switzerland ranks high in innovation, competitiveness and research spending, and much of this is enthusiasm for research is concentrated around Zurich. UZH also benefits from synergies with the ETH Zurich.

The Mathematics Institute has 17 professors and around 60 PhD students, part of a graduate school run jointly with ETH Zurich. Also joint is a Computational Science program uniting 47 researchers, mostly in the sciences, who make use of computational methods.

## Curriculum vitae

**Paul-Olivier Dehaye** Paul-Olivier Dehaye is a Swiss National Science Foundation Assistant Professor at the University of Zurich. After his PhD at Stanford (2006), he has also worked in Oxford, at the Institut des Hautes Etudes Scientifiques and at ETH Zurich. He currently has 13 papers published in international peer-reviewed journals. He is currently supervising three PhD students and one post-doc.

His main research is at the intersection of Number Theory and Combinatorics, and in particular in Random Matrix Theory conjectures. He has additional interests in FLOSS, semantic tools, massive online education and crowdsourcing, all with the view of enabling larger scale mathematical and scientific collaborations. He is also member of the program committee of CICM 2015 (Conference on Intelligent Computer Mathematics).

He is a contributor to the Sage, LMFDB and OpenEdX projects, and has organised two conferences relating to these projects. The first was held in 2013 in Edinburgh, and organised jointly with Nicolas Thiery. Its official title was *Online databases: from L-functions to combinatorics*, and it served as a precursor to some aspects of this grant, by bringing the Sage-Combinat and LMFDB communities together. The second was held in June 2014 in Zurich and organised jointly with Stanford. It aimed at building a community around the open source python-based MOOC platform OpenEdX, and opened a series of conferences now held twice annually.

Dehaye has also taught for two years now a python course using OpenEdX, which aims to bring first year students to the level of potential contributor to Sage. This course also has a project-based component. It is now run locally for a small audience, but could be scaled up in various ways.

## Publications, products, achievements

1. Dehaye is editor for the LMFDB, and has contributed to the project since its inception (2007). His students are also contributors.
2. For several of his papers, Dehaye used extensive computer-assisted experimentation (using mostly the combinatorial components of Sage) to inform the formulation of the eventual theorems, including for instance:
  - *Combinatorics of lower order terms in the moments conjecture for the Riemann zeta function*, arXiv:1201.4478
  - *Integrality of hook ratios*, arXiv:1111.5959, in *Proceedings of the Formal Power Series and Algebraic Combinatorics 2012 (Nagoya) conference*.
  - *A multiset hook length formula and some applications*, with Guoniu Han, in *Discrete Mathematics*, (311) 23–24, pp. 2690–2702, 2011.
  - *A note on moments of derivatives of characteristic polynomials*, in *DMTCS Proc. Formal Power Series and Algebraic Combinatorics 2010*, vol. 12.
  - *Joint moments of derivatives of characteristic polynomials*, in *Algebra and Number Theory Journal* **2** (2008), no. 1, pp. 31–68.
3. Dehaye has been extensively involved in teaching Python and Sage at UZH, through an online platform called OpenEdX. This has led him to organise the first community-driven conference around this (open-source) software, and to develop (together with students) additional tools, such as edx-presenter.



## Previous projects or activities

**Swiss National Foundation PP00P2/138906: Combinatorics of partitions and number theoretic aspects** This grant covers research at the intersection of number theory and combinatorics. Some of its aims are to uncover combinatorial structures that lurk in complicated formulae for moments of  $L$ -functions (such as the Riemann zeta function). As such, it is simultaneously a heavy user of numerical methods from analytic number theory and of combinatorial techniques implemented in Sage.

## Significant infrastructure

1. The Faculty of Sciences of the UZH benefits from very strong specialized IT support in the form of the S3IT group. They operate for instance a research cloud and a local supercomputer, and provide further assistance for the design of hardware and software systems to further research. They have a pool of software engineers that can be hired on projects such as this one for shorter periods.
2. UZH has a stake in Piz Daint, currently the sixth largest (and most energy-efficient) supercomputer in the world. This supercomputer is now currently expanded.

**University of XXX**

**Curriculum vitae**

**Publications, products, achievements**

1. ♠XXX [WRITE HERE: ...]♠

**Previous projects or activities**

1. ♠XXX [WRITE HERE: ...]♠

**Significant infrastructure**

♠XXX [WRITE HERE: ...]♠

## University of Southampton

### Curriculum vitae

#### Publications, products, achievements

1. ♠XXX [WRITE HERE: ...]♠

#### Previous projects or activities

1. ♠XXX [WRITE HERE: ...]♠

#### Significant infrastructure

♠XXX [WRITE HERE: ...]♠

♠Logilab [WRITE HERE: Description of Logilab]♠ ♠MK [WRITE HERE: Description of University of Silesia]♠  
♠LD [WRITE HERE: Description of Versailles]♠ ♠MK [WRITE HERE: PIC for University of Silesia]♠ ♠LD  
[WRITE HERE: PIC for Versailles]♠  
[?]

## 4.2 Third Parties Involved in the Project (including use of third party resources)

♠EC Commentary: *Please complete, for each participant, the table (see page 27 of "VRETemplate.PDF"), or simply state "No third parties involved", if applicable.*♠

No third parties involved.

♠TO DO: *Or Seattle?*♠

## 5 Ethics and Security

♣**EC Commentary:** *This section is not covered by the page limit.*♣

### 5.1 Ethics

♣**EC Commentary:** *If you have entered any ethics issues in the ethical issue table in the administrative proposal forms, you must:*

- *submit an ethics self-assessment, which:*

- *describes how the proposal meets the national legal and ethical requirements of the country or countries where the tasks raising ethical issues are to be carried out;*

- *explains in detail how you intend to address the issues in the ethical issues table, in particular as regards: research objectives (e.g. study of vulnerable populations, dual use, etc.), research methodology (e.g. clinical trials, involvement of children and related consent procedures, protection of any data collected, etc.), the potential impact of the research (e.g. dual use issues, environmental damage, stigmatisation of particular social groups, political or financial retaliation, benefit-sharing, malevolent use , etc.)*

- *provide the documents that you need under national law(if you already have them), e.g.:*

- *an ethics committee opinion;*

- *the document notifying activities raising ethical issues or authorising such activities*

*If these documents are not in English, you must also submit an English summary of them (containing, if available, the conclusions of the committee or authority concerned).*

*If you plan to request these documents specifically for the project you are proposing, your request must contain an explicit reference to the project title*♣

### 5.2 Security

Please indicate if your proposal will involve:

- activities or results raising security issues: NO
- 'EU-classified information' as background or results: NO

## References