H2020 Call 3 Topic 9-2015

e-Infrastructures for Virtual Research Environments (VRE)

EINFRA-9

OpenDreamKit: Open Digital Research Environments for Mathematics, as a Kit

Acronym: OpenDreamKit

site=PS
site=LL
site=UV
site=UJF
site=UB
site=UO
site=USH
site=USO
site=SA
site=UW
site=JU
site=UK
site=US
site=ZH
site=SR
site=UWS
acro=OpenDream
inst=H2020 Call 3 Topic 9-2015

H2020 Call 3 Topic 9-2015
Date of Preparation: January 5, 2015
Revision: of

Work program topics addressed by OpenDreamKit: Challenge TODO: N/A, **Objective N/A**: TODO, **target outcome N/A**) N/A.

Coordinator: Nicolas M. Thiéry

e-mail: Nicolas.Thiery@u-psud.fr

tel/fax: TODO

#	Participant organisation name	Short name	Country
1	Université Paris Sud	UPS	FR
2	Logilab	Logilab	FR
3	Université de Versailles Saint-Quentin	UVSQ	FR
4	Université Joseph Fourier	UJF	FR
5	Université Bordeaux	UB	FR
6	University of Oxford	UO	UK
7	University of Sheffield	USHEF	UK
8	University of Southampton	USO	UK
9	University of St Andrews	USTAN	UK
10	University of Warwick	UW	UK
11	Jacobs University Bremen	JacU	DE
12	University of Kaiserslautern	UK	DE
13	University of Silesia	US	PL
14	Universität Zürich	UZH	CH
15	Simula Research Laboratory	Simula	NO
16	University of Washington at Seattle	UWS	US

Abstract

♠**TO DO**: *expand (maximum: 2000 characters)*♠ OpenDreamKit focuses on delivering a toolkit that will make it easy to setup Virtual Research Environments, customized 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.

Contents

Outline of Project (for Proposers)

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

Vision

An internal attempt at specifying our vision through short (unsubstantiated) answers.

```
> 1) Who are we?

Lead or core developers of some of the major open source components
for pure mathematics and applications:

- Computational components: GAP, Linbox, MPIR, Pari, Sage, Singular
- Databases: LMFDB (findstat as well)
- Knowledge management: MathHub

Together with, in a larger scientific domain, lead developers for:

- Collaborative user interfaces (IPython, SageMathCloud)
- Database and Scientific Computing for the industry (Logilab)
- Numerical code optimization/parallelisation (Pythran)

> 2) What is our goal?

Building blocks with a sustainable development model that can be
seamlessly combined together to build versatile high performance
VRE's, each tailored to a specific need in pure mathematics and
application.

> 2.5) What is our strategy?

Maximize sustainability and impact by reusing and improving existing
building blocks, and reaching toward larger communities whenever possible.
E.g. factoring out our common user interface needs at the level
of IPython/Jupyter will save us time (sustainability), and impact
the larger scientific computing community.
The improvements to the building blocks will impact all their users,
whether they use the VRE or not.

> 3) From where do we start?

- Building blocks with a sustainable development model
- Proof-of-concept prototypes of VRE (SMC, Simulagora)
- Experience on combining together some of the building blocks

> 4) How do we connect or differ from other projects?

The other projects focus on either one or a few of the building
blocks, or on a specific VRE.

We articulate our work with each of them.

> 5) Why are we excellent?
```

```
The consortium puts together recognized experts in all
areas and most building blocks that are relevant to the goal. There is
simultaneously a variety of point of views and a record of past
experiences collaborating together at smaller scale
(e.g. GAP-Singular). The approach is bottom up.  Most joint tasks
consist in bringing together people with a common need. There is
experience in community building.  Most participants are
simultaneously users and developers of their tools.

All of this makes me confident that we will indeed be able to
productively collaborate. And do stuff that is first class and useful.

On Sat, Dec 13, 2014 at 11:18:10PM +0100, Wolfram Decker wrote:
> 0) What precisely is our starting point and why are we the right people to
> achieve what we promise to do? Are we leaders in the area touched
> by the proposal? How do we connect? Is there some past
> collaborative success?
> 1) You still do not say what we actually will provide. What precisely will
> the VRE offer to its users?

I more or less answered those points above. Let me know if I should
elaborate.

> Who will be its users? Will those already familiar with the involved
> CAS use it? Will it make the CAS more attractive for a much larger
> community?

One objective is definitely to make CAS and others more attractive by
lowering a lot the entry barrier to access the soft (and db, ...). A
typical situation that most of us ran into is, when collaborating with
other less tech-savvy mathematicians, to have trouble sharing code,
data, and in-the-writing papers with them. SMC was launched with this
idea in mind, and the success proves the concept.

At the same time, the improvements in the building blocks will also
impact CAS users that are happy with their current user interface /
work-flow.

Improvements to IPython will impact a much larger community.

> 2) You motivate what we wish to do by the success of SageMathCloud.
> But why do we than need another VRE? How do we differ from
> SageMathCloud?

There is no one-size-fits-all VRE. One might want to run a VRE on
one's own computer resources for a variety of reason (speed of access,
specific resources, privacy, independence, ...). One might want a
different combination of software (e.g. a lightweight VRE with only
Singular).  One might want to focus on data with LMFDB-style database
searches, or on interactive computing, or on document writing, or some
combination thereof.

> Do we have a chance to compete? Or will we rather join forces? In
```

```
> which way?

We join forces (the plan is to have William/UW in the consortium, as
non funded participant). SMC focuses on one specific cloud based
VRE. We focus on the building blocks and the glue. Both project are
mutually beneficial. See the language p. 14 of the proposal.

> 3) You motivate what we wish to do by the success of LMFDB. But what
> are our connections to this database? Will we enhance it? Will we connect
> it to other stuff we do? Will we create other databases?

LMFDB is a prototype of large scale database. We want to make it
easier for other groups of mathematicians to setup similar databases
in their area. Reciprocally, like SMC, the LMFDB with benefit back
from the improved building blocks.

> 4) Why is Europe in the lead if there is already SageMathCloud?
> Where precisely is Europe in the lead?

Europe is the lead in many of the building blocks.
```

Description of the call

```
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?)
```

B.0.1 Excellence

OpenDreamKit focuses on delivering tools that will make it easy to setup Virtual Research Environments, customized 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/GP`, `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.

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.

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

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

In particular, we will focus on developing and improving building blocks of e-infrastructure that can be assembled and re-used flexibly to address a wide range of requirements in mathematics and the applications of mathematics in science and engineering, rather than creating one particular monolithic environment.

B.0.1.1 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 OpenDreamKit are:

Aim 1: Improve the productivity of researchers in pure mathematics and applications 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, resources 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, producing reusable and easily accessible software, sharing data in a semantically sound way and exploiting and supporting the growing ecosystem of computational tools.

Aim 4: Maximize sustainability as well as impact in mathematics, neighboring fields, and scientific computing.

♠**TO DO**: *HF: Aim 3 seems rather crowded. Should we filter out 'reproducibility' and move this into Aim 4 (impact)? Or create a new aim?*♠

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 OpenDreamKit are:

Objective 1: To develop and standardize an architecture allowing a range of mathematical and data and software components to be combined with off-the-shelf non-mathematical infrastructure to produce specialized 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 or not suitable. These components should support VREs running on a variety of platforms, including standard cloud and cluster platforms. This primarily addresses Aim 2, thereby contributing to Aim 1 and 3.

♠SL [**WRITE HERE**: objective:core: Write detailed description]♠

Objective 3: To bring together related research communities (such as users of IPython, `Sage`, `Singular`, and `GAP`) to symbiotically exploit overlaps in tool creation building efforts, to avoid duplication of effort in different disciplines, and to benefit from best-practice pioneered in one community by sharing it most effectively. This supports Aims 1, 3 and 4.

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.

♠SL [**WRITE HERE**: objective:updates: 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 and 4.

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. ♠DP [**WRITE HERE**: objective:social write detailed description]♠ ♠SL [**WRITE HERE**: objective:social 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. ♠POD [**WRITE HERE**: write detailed description of Objective Data; improve concistency of the short description w.r.t. the Databases workpackage (in particular provenance w.r.t. tracability)]♠

Objective 8: Demonstrate the effectiveness of Virtual Research Environments built on top of OpenDreamKit 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 and through documenting best practice in reproducible demonstrator documents Aim 3.

Objective 9: Effective Dissemination ♠VP [**WRITE HERE**: objective:disseminate: 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/GP`, 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 programs 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 programs, 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 it's 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.

♠**TO DO**: *Keyword: flexible/versatile virtual environment. Does this belong here*♠

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 ??: Core components

Most of the direct mathematical capabilities of our software will come from existing or updated open source mathematical systems. For instance we will use the power of the GAP Library for computational group theory or `PARI/GP` 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 or adapted tools will be needed. One example is a general infrastructure for mathematical databases, covering some of the types of data values and search criteria common in mathematics, but rare outside, and issues such as provenance and citation that are common to most mathematical databases. Other examples include adapting user interface and collaboration tools to support mathematical notations.

Objective 3: Community building across disciplines

Open source tools are developed with great effect across different communities often leading to re-inventions and re-implementations of great ideas by different people not being aware of each other. By fostering a more cross-discipline community, by sharing tools where possible and by creating slightly more generic tools that can be shared, we will reduce such duplication of effort significantly. This will also lead to better quality software as the maintenance and development effort can be focused on one tool rather than a wide spread of codes, allow sharing of innovative ideas and practice more efficiently, and increasing productivity of research substantially.

While each of the communities such as `Sage`, `Singular`, and `GAP` need somewhat special features for their research, they are united through being (i) focussed on mathematical challenges, (ii) needing a computational workflow. `IPython` and the `Jupyter` Notebook are used widely in communities outside mathematics in Science and Engineering that also need the computational workflow and are based on (iii) applications of mathematics. These three common attributes are sufficient to distill the requirements for core features of the virtual research environment described in this OpenDreamKit proposal. Community building will also help to sustain ongoing and community driven maintenance of a such a tool.

Objective 4: Updates to Mathematical Software components Essential to our vision is leveraging the communities massive decades-long investment in a range of open source mathematical software systems. These systems are complex, widely used and powerful, but generally designed for operation as stand-alone programs, not as part of an integrated VRE. Many are also not well-suited for modern platforms, needing work to better support parallel programming, virtualization and HPC platforms. We will update these systems to work together seamlessly and comply with best practice for portability and platform integration.

Objective 5: A sustainable ecosystem of software components

The success of large specialized software like `PARI/GP`, `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 decade has proven the viability of the “developed by users for users” model: despite its large international 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 possible 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), by spinning off software development (e.g. the `Cython` compiler) to larger communities whenever possible, and by carefully designing the development workflow.

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 OpenDreamKit will invest toward, factoring out joint needs, and outsourcing or spinning off more components to larger communities. For example, OpenDreamKit 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.

OpenDreamKit 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 6: ♠TO DO: *TITLE: SOCIAL*♠ ♠DP/... [**WRITE HERE**: objective:social: detailed description]♠

Objective 7: Next generation mathematical databases ♠POD [**WRITE HERE**: objective:data: detailed description]♠

Objective 8: ♠TO DO: *TITLE: DEMO*♠ ♠ALL [**WRITE HERE**: objective:demo: detailed description]♠

We will create a number of demonstrators to highlight the power of OpenDreamKit, to show the applicability across many different domains in mathematics and science, to provide recipes for others and to document what is possible with state-of-the-art computational infrastructure tools.

In particular, we will create interactive books for three different application domains (engineering mechanics, biology and physics) demonstrating a variety of numerical and symbolic techniques in self-contained executable documents. These will be of great value as textbooks for university and high school education, for outreach activities and for self-study. We will further develop Notebook tools for a magnetic materials simulation package to demonstrate the value of OpenDreamKit for leading edge computational science, and develop the corresponding executable tutorials and documentation.

In addition, all four demonstrators will illustrate the features developed in this project, can be re-executed as a regression test, and form part of the documentation of OpenDreamKit.

Objective 9: ♠**TO DO:** ***Title: disseminate***♠ ♠VP [**WRITE HERE**: objective:disseminate: detailed description]♠

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/GP`, Singular, `Jupyter`, ...) 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, OpenDreamKit 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, and we will run a series of workshops to inject additional momentum into the process. Doing so, our objective is to help the communities to grow themselves and interact together using our work.

B.0.1.2 Relation to the Work Programme

OpenDreamKit addresses the topic “E-infrastructures for Virtual Research Environments (VRE)”, under E-Infrastructures-2015 call. In the table below we explain how this project addresses the specific challenge and the scope of that topic, as set out in the work program.

♠NT [**WRITE HERE**: Fix the margins for the table below]♠

Specific challenge	OpenDreamKit contribution
Empower researchers through development and deployment of service-driven digital research environments, services and tools tailored to their specific needs.	OpenDreamKit will empower researchers in mathematics and application veloping a service-driven tool, based on software, knowledge and data int Tailored to the researchers’ specific needs and workflows, the VRE will sup entire life-cycle of computational work in mathematical research. It will imp productivity within the community by investigating better collaboration proces identifying, sharing and promoting software development best practices.
VRE should integrate resources across all layers of the e-infrastructure (networking, computing, data, software, user interfaces)	OpenDreamKit will indeed integrate resources across all layers of the e-infra : software development models, collaborative tools, data, component arch deployment frameworks, standardization, social aspects, but also fostering ration inside the community, community enlargement and links with other communities.
VRE should foster cross-disciplinary data interoperability.	OpenDreamKit will foster a sustainable ecosystem of interoperable source nents developed by overlapping communities, and data interoperability betw ferent fields of mathematics.
VRE should provide functions allowing data citation and promoting data sharing and trust.	The project will allow an easy, safe and efficient storage, reuse and sharin mathematical data, taking account of provenance and citability. It will allow d ing in a semantically sound way, and make software sustainable, reusable a accessible.
Scope	OpenDreamKit contribution
Each VRE should abstract from the underlying e-infrastructures using standardized building blocks and workflows, well documented interfaces, in particular regarding APIs, and interoperable components	We will use building blocks with a sustainable development model that can b lessly combined together to build versatile high performance VREs, each tail specific need in pure mathematics and application. We will develop and dem (WP3) a set of APIs enabling components such as database interfaces, c tional modules, separate systems such as `GAP` or `Sage` to be flexibly comb run smoothly across a wide range of environments (cloud, local, server etc.). well defined APIs, we will enable discovery of subsystems, functionality, doc tion and computational resources.
The VRE proposals should clearly identify and build on requirements from real use cases	OpenDreamKit will be built on the requirements from the real use cases, those involving industrial stakeholders. At the end of the project, the effectiv the VRE will be demonstrated for a number of real use cases from different d
They should re-use tools and services from existing infrastructures and projects at national and/or European level as appropriate.	OpenDreamKit project brings together and integrates already existing tools a active scientific computing environments : `GAP`, `Sage`, `LinBox`, `PARI/GP`, `Si` and `IPython`, connected to databases, that will allow a huge gain in efficie productivity, enabling a large-scale collaboration on software, knowledge, an
Where data are concerned, projects will define the semantics, ontologies, the *what* metadata, as well as the best computing models and levels of abstraction (e.g. by means of open web services) to process the rich semantics at machine level, as to ensure interoperability.	We will investigate patterns to share data, ontologies, and semantics across tational systems, possibly connected remotely. We will leverage the well es semantics used in mathematics (categories, type systems) to give powerful tions on computational objects.

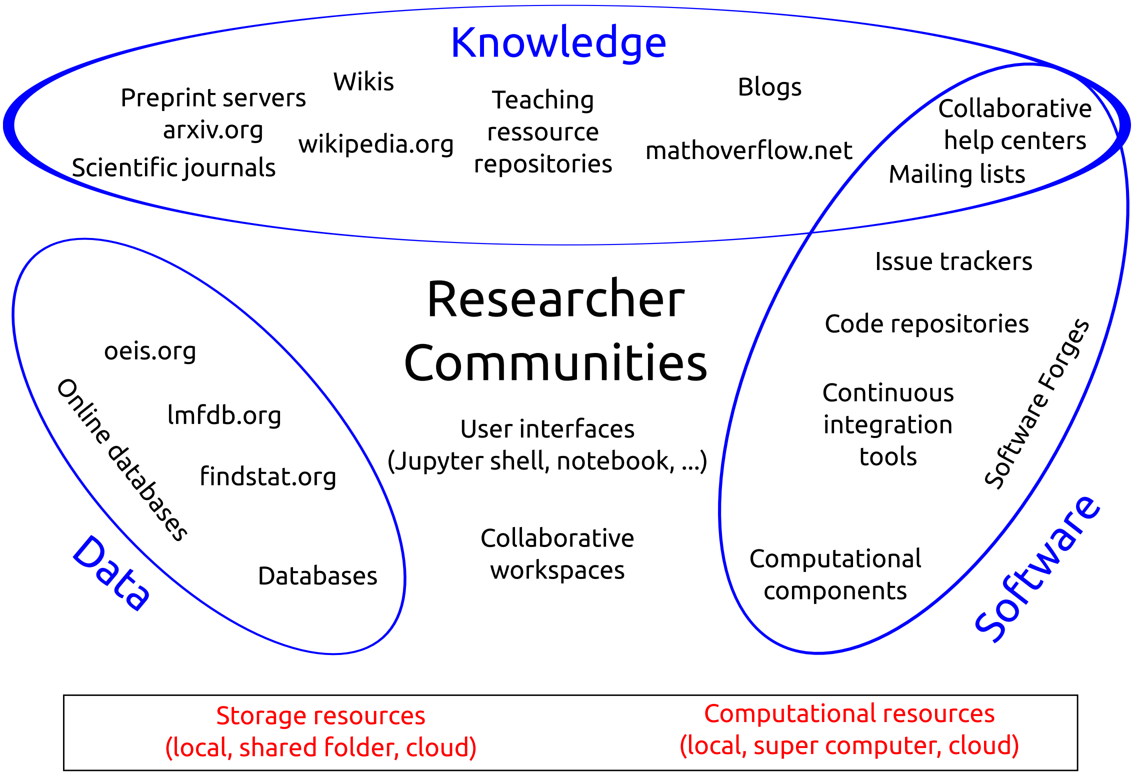

Figure 1: Virtual Research Environments for research in pure mathematics and applications.

B.0.1.3 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.♠

♠**TO DO**: *NT: the purpose of Figure 1 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*

Improvements to implement:

- *the findstat link does not work for me, kerning looks extremely weird – POD*

- *LMFDB, OEIS, and findstat have a strong knowledge component as well, with knowls and wikis, references, ...*
- *arxiv is not far from a database of knowledge*

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

♠

B.0.1.3.1 Importance of experimental tools in pure mathematics and applications

From their early days, computers have been used in pure mathematics, either to prove theorems (e.g. the four color theorem) or, like the telescope for astronomers, to explore new theories. By now the experimental method, based on exact computer aided calculations, has now been added to the standard toolbox of the pure mathematician, and its usage has grown to the point that certain areas of mathematics now completely depend on it.

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 Millenium 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 cross-connections between different areas of mathematics also computationally and, thus, to arrive at cutting-edge applications which previously were inconceivable.

The field of computational mathematics 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 (♠**TO DO**: *this is a bit redundant with below*♠), applications of the computational methods in various fields, and the creation of widely used databases.

Particularly 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.

B.0.1.3.2 A long track of collaboration on software, data, knowledge

Supporting the experimental method requires spending major efforts 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 specialized systems like `GAP`, `Singular`, or `PARI/GP` playing a major role for decades.

The next scale was reached in the last decade with the advent of the general purpose mathematical system `Sage` which proved the viability and sustainability of the "developed by users for users" development model at the international level.

♠**TO DO**: *This is somewhat redundant with the language in Objective 5; see where this belongs best to.*♠

♠**TO DO**: *Develop*♠Similarly, 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.

♠**TO DO**: *Develop*♠Mathematicians have a strong tradition of sharing knowledge openly (arxiv, Wikipedia, ...).

B.0.1.3.3 Early VRE's

♠**TO DO**: *Motivate the relevance of VRE's, in particular by the success of `SageMathCloud` or `Simulagora`. Mention as well `LMFDB`.*♠

♠**TO DO**: *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*♠

B.0.1.3.4 Key concept: bringing communities together toward a VRE kit

♠**TO DO**: *Focus on VRE kit and building blocks*♠

♠**TO DO**: *Why this focus? variability of needs, sustainability, ...*♠

♠**TO DO**: *Bringing communities together*♠

B.0.1.3.5 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 `computeralgebra.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 cross-linked 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.

The work in the SPP1489 has established effective communication channels between the core developers of different computer algebra systems. It is a showcase project for several objectives of this proposal (such as community building and fostering a sustainable ecosystem of interoperable open source components). The experience made in parallelizing mathematical software will be crucial for Work package WP5.

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. Sloanfoundation 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 OpenDreamKit, especially given the current rapid evolution of IPython toward its language agnostic avatar Jupyter.

OpenDreamKit 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.

NSF SI2-SSE OCI-1147247 The OCI-1147247 Collaborative Research grant "Sage-Combinat: Developing and Sharing Open Source Software for Algebraic Combinatorics" is a project funded by the National Science Foundation from June 2012 to May 2015. The grant supports the development of `Sage-Combinat`, on the USA side, and in areas relevant to the ongoing research of the participants (symmetric functions, Macdonald polynomials for arbitrary Cartan types, crystals, rigged configurations and combinatorial R-matrices, affine Weyl groups and Hecke algebras, cluster algebras, posets, ...), together with relevant underlying infrastructure. The grant funds a yearly Sage Days workshop, and cofunded two others at ICERM and Orsay respectively. The grant also funds a dedicated software development and computation server for `Sage-Combinat`, hosted in the `Sage` computation farm in Seattle. Emphasis is placed on the development of thematic tutorials that make the code accessible to new users. The grant also funds graduate student RA support, curriculum development, and other mentoring.

Two of the proposers, Stein and Thiéry, are respectively PI and foreign senior participant to this NSF grant. It funded, through them, some of the development of `SageMathCloud` as well as of the category framework in `Sage`; both are key assets for this proposal. The workshop and outreach actions pursued by this NSF grant have proven to be potent tools for connecting researchers and recruiting users and developers. One of the role of this proposal is to support similar community building in Europe.

HPAC grant from the A.N.R. The French national research agency ANR has funded a 4 years project on High Performance Algebraic Computing (HPAC) focused on the development of parallel exact linear algebra. The consortium gathers research groups from LIP6 (Paris 6), LIRMM (Montpellier), LIP (Lyon) and LIG and LJK (Grenoble). The main goals of the project is to first develop high performance exact linear algebra kernels with dedicated parallel runtime, propose a domain specific language for the parallelization of exact linear algebra libraries and their composition, invent new algorithmic solutions for large scale parallelizations. The output of the project is then twofolds: new computational challenges arising in algebraic cryptanalysis will be addressed, and the open-source libraries maintained by each group will not only integrate these advances, but will expose them in a close integration to high level computer algebra softwares. In this process, `Sage` will start benefitting from the new shared-memory parallel code of `LinBox` for the linear algebra over a finite field. The scope of this project is mostly focused on shared memory parallelism (except for some challenge computations). Addressing distributed and heterogeneous infrastructures is the next step after this project, that is be addressed in work-package 5 of the this proposal.

RADIANT Grant from EU FP7-HEALTH (ref 305636) `http://radiant-project.eu/`

♠**TO DO**: *Neil: This is a possible inclusion I think it could be a good archetype for how we help science in general, but please remove if it's inappropriate!*♠

This EU funded proposal focuses on making available computational and mathematical models to the computational biology communities as rapidly as they are developed with a particular focus on high throughput sequencing techniques. The rapid development of sensorics technology in the biological sciences results in mathematical challenges in the data analysis. To address these challenges in a timely manner collaborative frameworks for mathematical and computational modelling are required. The DreamKit project provides the framework for pipeline delivery of methodologies to end users through approachable IPython/Jupyter notebooks.

Logilab: simulagora, cubicweb, ...

RADIANT Grant from EU FP7-HEALTH (ref 305636) `http://radiant-project.eu/`

♠**TO DO**: *Neil: This is a possible inclusion I think it could be a good archetype for how we help science in general, but please remove if it's inappropriate!*♠

This EU funded proposal focuses on making available computational and mathematical models to the computational biology communities as rapidly as they are developed with a particular focus on high throughput sequencing techniques. The rapid development of sensorics technology in the biological sciences results in mathematical challenges in the data analysis. To address these challenges in a timely manner collaborative frameworks for mathematical and computational modelling are required. The DreamKit project provides

the framework for pipeline delivery of methodologies to end users through approachable IPython/Jupyter notebooks.

Logilab: simulagora, cubicweb, ...

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

Sage Math Cloud https://cloud.sagemath.com/

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 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 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 open-source cloud-based Virtual Research and Teaching Environment for mathematics developed since 2013 under the lead of William Stein, with funding from the NSF, and Google's Education Grant program. It's currently deployed at the University of Washington at Seattle, with a business plan in the work for commercial support for massive on line courses, subsidizing a free service for all other academic usage and some further Sage development.

In comparison OpenDreamKit 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 OpenDreamKit, paving the way and showcasing important features from the users perspective. On the other hand, basically each and every task undertaken in OpenDreamKit 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, 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 theLMFDB 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). 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 OpenDreamKit as it would connect the project with a pool of experts. Joint workshops between the LMFDB and OpenDreamKit 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 OpenDreamKit 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

B.0.1.4 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.♠

For most pure mathematicians using computational tools in their research, the state of the art in 2014 is still a collection of programs each of which must be installed individually on their desktop or laptop computer, respecting a complicated graph of dependencies. Alternatively software may be installed on a departmental server or cluster and used via text-based remote login. The software performs computations (using excellent implementations of extremely sophisticated algorithms) with inputs and outputs usually in a bespoke text-based format. The results of computations are incorporated into publications by cut-and-paste and collaboration is through exchange of programs and data by email, shared general-purpose file servers or, rarely, a service such as GitHub. Multiple computations involved in producing a mathematical result must be managed by editing, naming and filing multiple scripts or programs, and there is no automatic support for rerunning computations to check for human or algorithmic error.

♠**TO DO**: *Work out best order for these topics*♠

There are commercial "symbolic computation systems" such as Mathematica ®. or Maple ®. which offer somewhat more modern frameworks, but they lack the specialised algorithms for research work in fields such as algebra, number theory or algebraic geometry and are not well-suited to support them.

The need for a more modern, more productive and less error-prone environment for this kind of mathematical research computing is widely acknowledged, but the separate groups developing the systems have individually, neither the time nor the expertise to develop it. There have been a number of interesting projects which have explored different aspects of what is needed: ♠**TO DO**: *SCIEnce, SageMathCloud, MathOverflow, Polymath, SAGE itself, SAGE notebooks, HPCGAP, recomputation*♠ and we will build on the experiences, and where useful the software, of all of these.

Our ambitious plan in this project is to learn from, and leapfrog, these piecemeal developments and provide a toolkit of software and interfaces, which supports the whole mathematical research process in a way which is **modern**, **seamless**, **collaborative**, mathematically **rigourous** and **adaptable** to the diverse needs of different mathematical research areas and of different mathematicians and collaborations.

♠**TO DO**: *Explain all the highlighted words in the para above*♠

♠**TO DO**: *Some examples here – what will we deliver to individuals; small "single-problem" collaborations; longer-lived data- or algorithm- centred teams; massive "flagshsip" projects. Think "User Stories"*♠

B.0.1.4.1 Challenges specific to mathematics

Mathematical research, especially pure mathematics presents some unique challenges to the realisation of this ambition.

♠**TO DO**: *Evidence this in more detail, clean up the language*♠

- Community mainly made of individuals or *very* small groups (1 PI + a few students). Few formal or structured research groups such as you might find in an equipment-intensive science. Large scale collaborations happen (CoFSG, Polymath), but still driven by individuals, not formal structures or money.

- Much top quality research has little or no formal research funding. So computational resources are limited to what is available anyway – personal laptops, departmental clusters...

- Many mathematical computations are highly irregular and complex. Traditional HPC paradigms coming from simulation and linear algebra do not apply.

- Mathematical notations have been refined over many centuries for use by humans with pen, paper and blackboard. Even such simple problems as selecting a sub-expression are hard to handle well on a computer. For instance $a + c$ is naturally seen as a subexpression of $a + b + c$ by a human.
- The mathematical correctness of widely used algorithms hinges on quite complex chains of reasoning. Subtle coding errors may easily produce plausible, but wrong, answers.
- Mathematical data different in several ways from typical scientific data
  - More often than not data is the result of a computation (and not a measurement of the real world). 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 ♠**TO DO**: *Someone explain this*♠
  - interlinking very high
  - several alternate and defining description of same objects

B.0.1.4.2 Challenges of a community built around multiple existing software projects

Another source of unique challenges for this project is the need to interact with several large and diverse ecosystems of software developers. For instance the `GAP` package development community, the SAGE development community, the wider Python community, the developers of key open-source libraries on which we rely and so on.

These communities exist in a delicate balance between collaboration and competition. For instance the SCIEnce project and SAGE were simultaneously exploring two different approaches to linking open-source mathematical software. Many technical developments (better IO handling in `GAP`, for instance) could usefully be shared, and at the end of the day we all want to do better mathematics, but a certain degree of competition is both natural and healthy.

In this project we need to build a sustainable “meta-ecosystem” in which systems may compete to have the best designs or algorithms, but all agree to cooperate on interfaces, bug reporting, testing, etc. to keep the final user experience seamless and reliable.

Promoting collaboration over competition between communities.

B.0.2 Impact

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

B.0.2.1 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.)♠

B.0.2.1.1 Impacts as listed in the work programme

The following table shows how OpenDreamKit addresses the specific impacts listed in the work programme.

More effective collaborations between researchers	
OpenDreamKit will strengthen collaborations between European scientific community in different branches of mathematics, through a development of the common e-infrastructure by bringing together software and databases previously separated, and build links with scientific communities in other disciplines (biology, physics, astronomy) that will use this e-infrastructure. The scientific community will considerably increase, by integrating new actors both from academic and non academic sector.	Key performance indicators : 1. By the end of the project, the scientific community in mathematics using the tool will increase by X % 2. The scientific community in other disciplines (biology, physics) using the tool will increase by X% 3. Number of enterprises using the tool will increase by X%
Higher efficiency and creativity in research, higher productivity of researchers thanks to reliable and easy access to discovery, access and re-use of data	
The development of a new integrated tool, replacing 3 previously separated tools, a possibility of real time data sharing, data re-use and simultaneous working will allow an important gain in time and in efficiency, and, by consequence, higher productivity of researchers. Moreover, the exchange of best practice (such as regular audit of codes) will have an important impact on the quality of the research. Finally, the unique tool will allow considerably reducing the costs of further research.	Key performance indicators : 1. The e-infrastructure developed by the end of the project will help reducing time of research by X% 2. The productivity of researchers will increase by X% 3. Thank to best practice exchange, the quality of research will considerably increase and number of errors will be reduced 4. Better traceability of research 5. Costs for research considerably reduced

Accelerated innovation in research via an integrated access to digital research resources, tools and services across disciplines and user communities	
An integrated access to digital research resources and tools that OpenDreamKit will provide will clearly help accelerating innovation in research across disciplines and communities. The integrated tool will meet the needs and help overcoming the obstacles that are common to several disciplines impacted by the project, and both to academic and non-academic research. Industrial actors, actively involved into the project will directly benefit from the project results.	Key performance indicators : 1. Emergence of new, improved research methods in several disciplines using the tool 2. Resolution of series of problems proper to industrial actors in several disciplines
Researchers able to process structured and qualitative data in virtual and/or ubiquitous workspaces	
OpenDreamKit will enable researchers to process different kind of data thanks to an integrated tool interconnected with databases. Efficient data storage will allow further exploitation and re-use of mathematics data for further calculations and thus make data processing more efficient. Le travail sur la sémantique de données est fait plus en amont.	Key performance indicators : Note E.S. – Nicolas, ici, j'ai du mal, car pas sure si on répond vraiment à ce critère.
Increased take-up of collaborative research and data sharing by new disciplines, research communities and institutions	
The project will clearly enhance a take-up of collaborative research and data sharing by and between new disciplines and research communities. The communities already using the parts of the tool, will be enlarged, involving more and more industrial actors and young scientists. By developing the tool, it will reinforce collaborations between different branches of mathematics (both pure and applied). Once developed, the tool will be opened for research in various disciplines, such as biology, physics, astronomy etc.	Key performance indicators : 1. Increased collaborations and data sharing between different communities in mathematics 2. Increased collaborations and data sharing between research communities in other disciplines 3. Increased number of users, enlarged research communities

B.0.2.1.2 Improving innovation capacity and the integration of new knowledge

Innovations developed by OpenDreamKit project will meet the needs of the following ecosystem participants:

1. Device / module vendors: hardware manufacturers, equipment manufacturers of smartphones, tablets, laptops;
2. Network providers: service providers, network infrastructure vendors (Avaya, Juniper, Extreme, Cisco. . .)
3. Platform providers
4. Cloud service providers: Software-as-a-Service, Platform-as-Service, Infrastructure-as-a-Service
5. Systeme integrators: end-to-end integration services and value-added services (Accenture, HP, IBM. . .)
6. End users: research communities; IT, healthcare, education and aeronautics stakeholders

Industrial stakeholders will be directly involved into the project and the VRE development, so that the tool will be exactly tailored to their specific needs – that are the same that the scientific community ones. Moreover, this will allow an early time-to-market and will facilitate the technology uptake.

In the table below we have specified different market needs, and the ways the project will address each of them :

Table XX

Needs of markets	How the project will address these needs
Performance gain	The tool will enable its users to combine functionality from XXX other mathematical software programs and programming languages -mainstream programming language Python? -which compilers?
Infrastructure capacity: newly built infrastructures with fast broadband connections are well positioned for adoption	OpenDreamKit will allow different groups of users to work simultaneously, and thus provide a considerably gain in efficiency.
Lower cost of scaling	An open source architecture brings affordability: people and organizations donate towards common goals, and small organizations can gain access to equipment and research talent typically only afforded by the largest firms. Resources integration will reduce considerably the time and the costs of operations.
Going beyond limitations of interconnect technology	An open source architecture brings creativity (=the best minds to solve a problem)
Enabling new applications and features	Through a series of links that will be created between previously separated tools, and data interoperability, the VRE will enable new applications and features
Early time-to-market (TTM): companies are looking for solutions that would improve the speed at which they can procure services to bypass traditional information technology departments	The speed of development will improve tremendously thanks to open source. The implication of industrial stakeholders into the development will allow to deliver a tool the suits the best to their needs, and thus speed-up the time to market and technology uptake.
Easy-to-use service: first-time experiences is crucial to gain acceptance	-Design? Ergonomie? Nice multiuser web-based graphical user interface -integration with learning tools? (example : interactive whiteboard) ♠**TO DO**: *E.S: Nicolas, il faut réflechir à la façon de rendre l'outil plus attractif pour la jeune génération (=future génération des chercheurs), car c'est la génération Y qui va définir des standards pour la suite - donc plus d'impact et à plus long terme. Réfléchis à la possibilité de rendre l'outil accessible sur mobiles, tablettes etc., mais aussi à la façon " intéressante " de l'enseigner.*♠

B.0.2.1.3 Other impacts (environmental and socially important impacts)

♠**TO DO**: *Neil: Can we not be more ambitious than simply replacing the other alternative? By working in an open way we should be able to achieve much more rapid deployment of ideas (like in the RADIANT project I mentioned above).*♠

Here, our mission statement will be the same as for `Sage`: to provide a viable free open source alternative to `Magma`, `Maple`, `Mathematica` and `Matlab`.

To make OpenDreamKit a new reference, it is important to focus on the young generation, which will be the future users.

Generation Y phenomenon:

1. generation Y accounts for 30% of the total projected population in 2025
2. key influencers to change in workplace habits
3. easy adaptability to technology -> collaborative work would make them more productive
4. compulsion to check smartphones for emails, texts or social media updates -> adoption of social networks.

2.1.4. Potential barriers and 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.)

The project does not arise any specific regulation issues; in the unlikely event that new norms would appear during the project, appropriate measures will be taken by the advisory board following advice of relevant experts and standards agencies on national and EU level.

B.0.2.2 Measures to Maximise Impact

B.0.2.2.1 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.♠

Three types of impact are possible with our dissemination and communication activities: (1) people or organizations are informed about OpenDreamKit; (2) people or organizations act and use our conclusions or results; (3) people or organizations contribute and help to develop or improve the research infrastructure. The second form of impact supposes that parties understand the messages. The third form supposes learning, which is a very high level of impact. In the following table, we have listed how the proposed measures will help to achieve the expected impact among our stakeholders and target audiences.

Table X. Draft 'plan for the dissemination and exploitation of the project's results'.

Target users	Measures during the project	Expected impact		
		(1)	(2)	(3)
Scientific community in mathematics (experienced researchers and PhD students)	1. Recruitment for the project of specialists from industrial sector and PhD students that are already a part of the community ; 2. 10 technical workshops organized in the frame of OpenDreamKit, 3. 10 scientific trainings/year, 4. Participation to FPSac, Isaac international conferences(4), international congress of mathematical software (1) 5. Workshop " Sage and women" in USA (1?) 6. 10 publications in (social aspects) 7. Software demonstration during the conferences>publication 8. 100 PhD students trained and accessing the infrastructure 9. Workshops for PhD students in Africa	X X X X	X X X X X X X	X X X X X X

Scientific community in other disciplines	1. Direct implication of the representatives of those disciplines into the project; 2. Annual participation in Pycon international conference 3. X scientific trainings to other communities; 4. News in Nature and other editions in other disciplines (specify!) 5. Up to X PhD students trained on the tool in biology, physics etc. 6. Workshop " sage & women " in USA> pour IPython	X	X X X X X X	X X X
Policy makers	Are not directly concerned by the tool, but can be informed via international conferences and publications	X		
Industry	1. Industrial stakeholders have common needs with academic researchers. They bring to the project their specific competences and human resources. They are actively involved into the workshops and trainings (50% of audience). The project aims the enlargement of the community thanks notably to new industrial actors (including in other disciplines and sectors). They will appropriate the tool by their direct involvement into the project, by participation to the workshops, trainings and conferences or by their usual information channels. 2. Annual participation in Pycon international conference 3. Certification by technology clusters	X	X X X	X
Standartisation agencies	1. At the end of the project, after internal standartisation, the new norms will be accorded with specialized agencies at national and EU levels		X	
Students	1. 5 MooCs destined to master students. 2. The tool will be used for the elaboration of pedagogical documents, referenced on the specific website 3. The projects results will be integrated into Master courses, and into teacher training courses	X X X	X X	
Civil society	1. Results will be presented on the annual event "Worldwide meetings of the free software". This event generally touches upon all free phenomena in the society, and involved various stakeholders, including civil society actors.	X	X	

Public at large	1. Series of actions in high schools led by PhD students to raise awareness of pupils, and especially girls, on mathematics research and scientific careers 2. Communication large public via annual events like " Science holiday " etc. 3. Vulgarization papers and communication events addressed to people interested by ICT. 4. Social networks and platforms	X X X X		

Measures after the project:

1. Continue dissemination to scientific community and industrial stakeholders through participation to international conferences (FPSac, Isaac, Pycon, Sage and Women etc.) and publications
2. Software demonstration during the conferences
3. Training of PhD students in mathematics, informatics and other disciplines, both in Europe and all over the world
4. Large public communication through regular vulgarization events

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).

Long term sustainability By design (Objective 1), the VRE's promoted by OpenDreamKit 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.

The needs in financial support after the end of the project are therefore not very important. We expect that a part of developers positions will be made durable by the partners institutions, thanks to an increase of awareness among them on the necessity of this infrastructure for their own needs.

With the increase of the number of users, more and more research laboratories, teaching institutions, and enterprises will get a need for using this VRE – thus, additional funding will be possible through access provision to other scientific communities, on projects base, or via service delivery.

♠**TO DO**: *None of the two models below really match our situation; investigate for some good picture and language, e.g. in "Economie du Logiciel Libre" of François Élie*♠

Here we propose two possible models of this use:

1. The **GPL model** (see Figure 2): With this model, the vendor is required to make the new code available in source form but it can choose to keep the new code as proprietary and charge for that proprietary software. The vendor can provide the code commercially as part of a larger platform (hardware/software product) for which the companies receives revenue (license fee for the code + fees for technical support, updates and upgrades).

2. The **Third Party Service Model** (see Figure 3) Many users may be willing to employ a third party service for distribution, modifications (debugging) and other support.

♠**EC Commentary**: *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?*♠

♠**TO DO**: *E.S. – Nicolas, ici je ne peux pas écrire à ta place. Il faut juste que tu répondes précisément aux questions posées ci-dessus.*♠

Open source software.

All software used and/or generated by the project will be Open Source. This is a deliberate choice of the project consortium, as commercial licenses (and patents) on this type of software only creates barriers in our scientific domain.

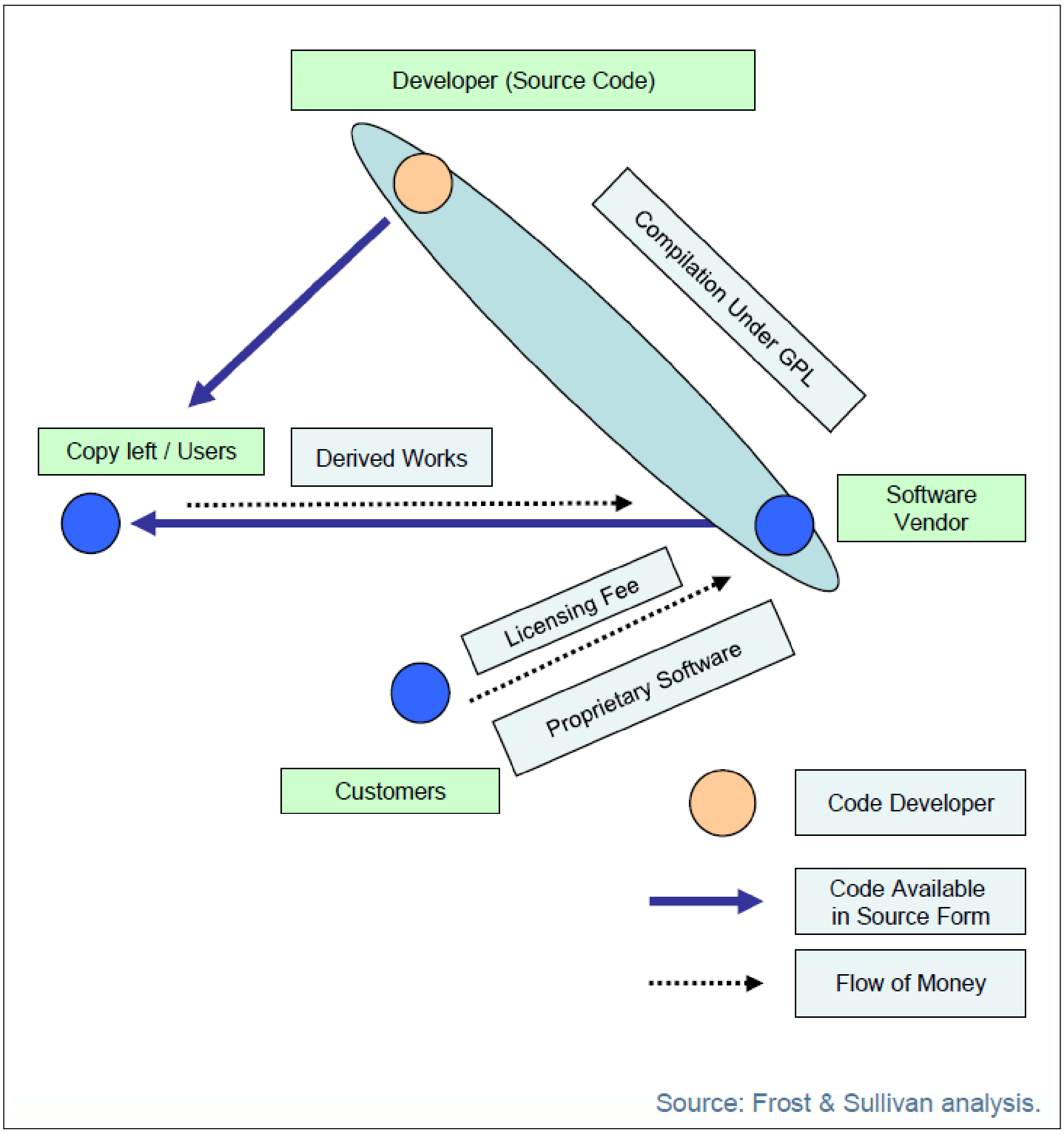

Figure 2: The GPL Model

Benefits of Open Source:

Acquisition and Costs: lower costs, easy access to the infrastructure, lower risks of proprietary lock-in

Flexibility: picking up from Open Source projects, reduces dependence on supplier, ability to view and modify the source code. Allows peer reviewed modifications, community discussions. Open Source provides the customer/end user the opportunity to innovate

Support: from developer community.

Besides being cost effective, Open Source software fosters reuse, reliability, flexibility, and interoperability.

A consortium agreement will be established to manage ownership and access to key knowledge, including software generated by the project.

Open access policy and data protection.

In line with the Horizon 2020 rules and current trends in IP management and publication strategies, our consortium is committed to open access publishing. This means that an article is immediately provided in open access mode by the scientific publisher. As the associated costs are usually shifted away from the

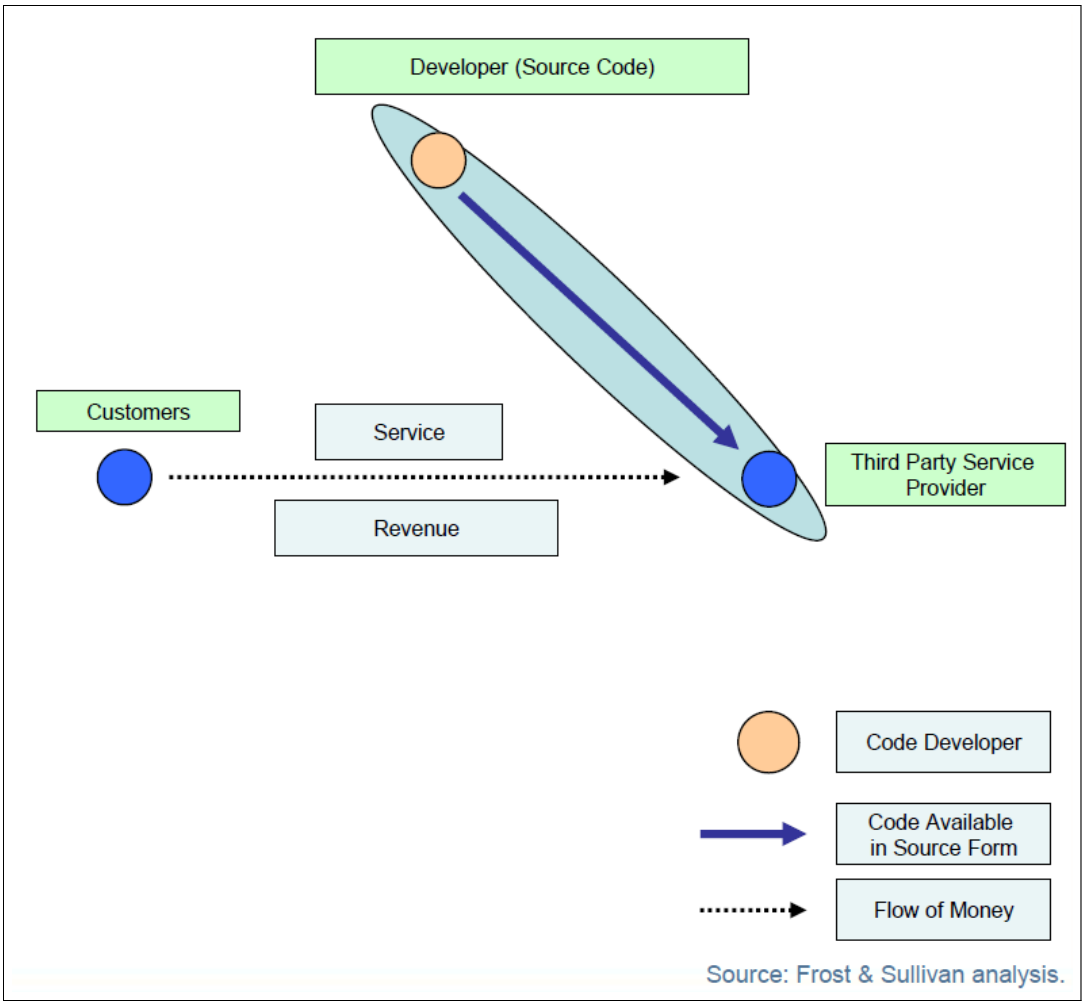

Figure 3: The Third-Party Services Model

readers, and instead to the university or research institute to which the researcher is affiliated, such costs have been accounted for by all the research partner institutions. We have agreed to follow the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. In fact, following the now well established tradition of the community, preprints for most publications will be posted on `arXiv`.

Conforming to the call requirement, we will also participate in Open research data pilot. Following the Guidelines on Data Management in Horizon 2020, we will establish a Data Management Plan, which first version will be provided as an early deliverable in first six months of the project. More developed versions of the plan will be provided as additional deliverables at later stages.

Our strategy for knowledge management and protection includes measures to provide open access ('green' or 'gold') to peer-reviewed scientific publications and all data that may result from the project.

B.0.2.2.2 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.*♠

B.0.3 Implementation

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

B.0.3.1 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 zeroork packages as shown in Figure 5: WP2 deals with ... In addition, there is one management work package (WP1) and one general dissemination work package (WP8). The Gantt chart on Page 29 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 B.0.1.1, page 3) 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 Plan Timing: GANTT Chart showing Task Dependencies and Information Flows

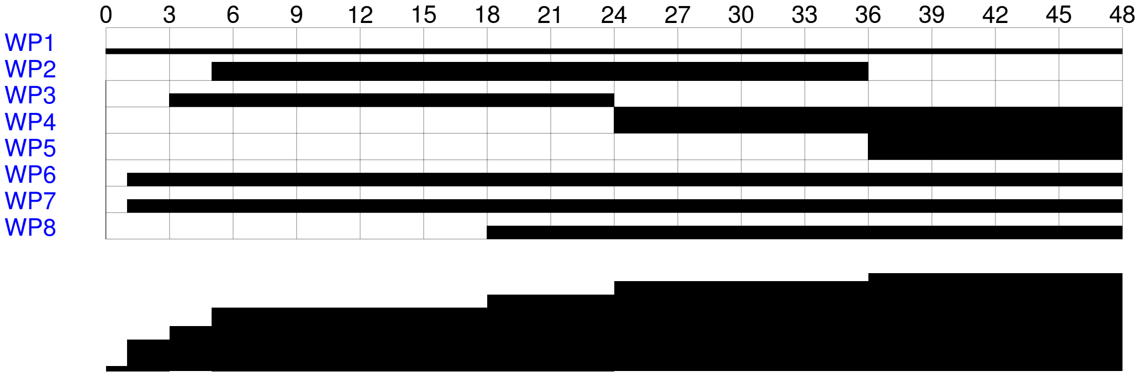

Figure 4: Overview Work Package Activities (lower bar shows the overall effort per month)

WP	Title	UPS	Logilab	UVSQ	UJF	UB	UO	USHEF	USO	USTAN	UW	JacU	UK	US	UZH	Simula	UWS	total
WP1	Management	*48*							2	48								**98**
WP2	Community Building and Engagement	*12*						8		1								**21**
WP3	Component Architecture	24		*1*				4		1								**30**
WP4	User Interfaces	1	1			1		12	21	1		12	1					**50**
WP5	High Performance Computing	1	12		12	1				1			1					**28**
WP6	Data/Knowledge/Software-Bases							12			3	*36*			1			**52**
WP7	Social Aspects						*1*	8	5									**14**
WP8	Dissemination							8	10	1				24				**43**
totals		**86**	**13**	**1**	**12**	**2**	**1**	**52**	**38**	**53**	**3**	**48**	**2**	**24**	**1**	**0**	**0**	**336**
intended totals		**48**	**48**	**48**	**48**	**48**	**48**	**48**	**48**	**48**	**3**	**48**	**48**	**48**	**48**	**48**	**48**	
requested totals		24	0	0	30	18	28		38	0	3	48	0	12	6	24	0	

Efforts in PM; WP lead efforts light gray italicised

Figure 5: Work Packages

Deliverables

#	Deliverable name	WP	Nature	Level	Due
M1.1	Create tickets for all relevant tasks / deliverables	WP1	DEC	PU	1
M1.2	Internal and external mailing lists	WP1	DEC	PU	1
M1.3	Internal software repository	WP1	DEC	PU	1
M5.18	Ensure interoperability of `Pythran` with `Python` and its packages.	WP5	R	PU	1
M5.17	Facility to compile `Pythran` compliant user kernels.	WP5	DEM	PU	2
M5.14	Turn the Python prototypes for tree exploration into production code, integrate to `Sage`.	WP5	DEM	PU	3
M5.16	Improve `Pythran` runtime support to automatically take advantage of multi-cores and SIMD instruction units.	WP5	DEM	PU	3
M5.4	Parallelise the relation sieving component of the Quadratic Sieve	WP5	DEM	PU	3
M2.1	Workshop 1	WP2	O	PU	6
M5.15	Implement `Pythran` runtime support in `Cython` when they are implemented instead of using default implementation.	WP5	DEM	PU	6
M4.9	Python Interface to OOMMF	WP4	O	PU	9
M5.5	Implement a parallel version of Block-Wiederman linear algebra over GF2 and the triple large prime variant.	WP5	DEM	PU	9
M6.13	Heuristic Parser for the OEIS	WP6	O	PU	9
M6.3	Workshop Report	WP6	R	PU	9
M1.4	Project Periodic Report (first year)	WP1	R	PU	12
M2.2	Report on community needs	WP2	R	PU	12
M3.1	One-click install `Sage` distribution for Windows with Cygwin 32bits	WP3	DEM	PU	12
M3.6	Virtual images and containers	WP3	O	PU	12
M4.1	Basic Jupyter interface for GAP, Pari, Sage, Singular	WP4	O	PU	12
M4.2	Full featured Jupyter interface for GAP, Pari, Singular	WP4	O	PU	12
M4.3	Sage notebook / Jupyter notebook convergence	WP4	DEM	PU	12
M4.5	Tools for collaborating on notebooks via version-control	WP4	O	PU	12
M5.11	Provide FAT binary support for all new x86_64 processors, build system maintenance and improvements to tuning, profiling and testing subsystems	WP5	DEM	PU	12
M5.14	Explore the possibility to interface smoothly `Pythran`, `Cython` and `Cilk++`.	WP5	DEM	PU	12
M5.19	Make `Pythran` typing better to improve error information.	WP5	DEM	PU	12
M5.1	Library design and domain specific language exposing `LinBox` parallel features to `Sage`	WP5	R	PU	12
M5.6	Take advantage of multiple cores in the matrix Fourier Algorithm component of the FFT for integer and polynomial arithmetic,and include assembly primitives for SIMD processor instructions (AVX, Knight's Bridge, etc.), especially in the FFT butterflies	WP5	DEM	PU	12
M5.7	Parallelise the Singular sparse polynomial multiplication algorithms	WP5	DEM	PU	12
M5.8	Parallel versions of the Singular sparse polynomial division and GCD algorithms.	WP5	DEM	PU	12
M5.9	Extend the existing assembly superoptimiser for AVX and upcoming Intel processor extensions	WP5	DEM	PU	12

M6.1	Conversion of existing and new databases to unified interoperable system	WP6	DEC	PU	12
M6.4	Design of Triform (DKS) Theories (Specification/RNC Schema/Examples)	WP6	R	PU	12
M6.6	Python/Sage Syntax Foundation Module in OMDoc/MMT	WP6	DEC	PU	12
M6.9	LMFDB deep modelling: Fragment Identification & Initial Model Design	WP6	R	PU	12
M7.1	...	WP7	??	PU	12
M8.3	Demo: Problems in Physics with Sage v1	WP8	DEM	PU	12
M4.10	Jupyter notebook Interface for OOMMF (OOMMF-NB)	WP4	DEM	PU	15
M2.3	Workshop 2	WP2	O	PU	18
M4.4	Facilities for running notebooks as verification tests	WP4	O	PU	18
M6.10	LMFDB Data vs. Knowledge vs. Software Validation	WP6	R	PU	18
M6.14	Cross-Validation for OEIS DKS-Theories	WP6	R	PU	18
M4.11	OOMMF-NB executable tutorial and documentation	WP4	DEC	PU	21
M1.5	Project Periodic Report (second year)	WP1	R	PU	24
M3.10	Integration between SageMathCloud and Sage's TRAC server	WP3	O	PU	24
M3.2	One-click install Sage distribution for Windows with Cygwin 64bits	WP3	DEM	PU	24
M3.4	Support for the SCSCP interface protocol in all relevant components (Sage, GAP, etc.) distribution	WP3	O	PU	24
M3.7	Modularization of the Sage distribution	WP3	O	PU	24
M4.12	OOMMF-NB dynamic web service available	WP4	DEC	PU	24
M4.13	OOMMF-NB virtual machine images available for download	WP4	O	PU	24
M4.6	Jupyter extension for 3d CFD visualization	WP4	O	PU	24
M5.10	Ongoing support of Intel, AMD, ARM, Sparc processors and new Operating System versions	WP5	DEM	PU	24
M5.14	Refactor the existing combinatorics Sage code using the new developed Pythran and Cython features.	WP5	DEM	PU	24
M5.2	Algorithms and implementations. Library maintenance and close integration in mathematical software	WP5	DEM	PU	24
M6.2	Shared persistent memoization library for Python/Sage	WP6	O	PU	24
M6.5	Implementation of Triform Theories in the MMT API.	WP6	O	PU	24
M6.7	Python/Sage Computational Foundation Module in OMDoc/MMT	WP6	O	PU	24
M2.4	Workshop 3	WP2	O	PU	30
M8.4	Demo: Problems in Physics with Sage v2	WP8	DEM	PU	30
M8.5	OOMMF-NB public repository and continuous integration	WP8	DEC	PU	32
M1.6	Project Periodic Report (third year)	WP1	R	PU	36
M3.3	Continuous integration platform for multi-platform build/test.	WP3	P	PP	36
M4.7	Exploratory support for live notebook collaboration	WP4	O	PU	36
M4.8	Exploratory support for semantic-aware interactive widgets providing views on objects represented and or in databases	WP4	DEM	PU	36
M5.3	Implementations of algorithms using distributed memory et heterogenous architectures: clusters and accelerators	WP5	DEM	PU	36
M6.11	LMFDB Algorithm verification wrt. a Triformal theory	WP6	O	PU	36
M6.8	Python/Sage Declarative Semantics in OMDoc/MMT	WP6	O	PU	36

M8.2	Demonstrator: Nonlinear Processes in Biology interactive book	WP8	DEM	PU	36
M8.6	OOMMF-NB publication manuscript completed	WP8	R	PU	36
M8.2	Demo: Classical Mechanics interactive book	WP8	DEM	PU	40
M2.5	Workshop 4	WP2	O	PU	42
M8.8	OOMMF-NB workshops delivered	WP8	O	PU	42
M8.7	Demo: Problems in Physics with Sage v3	WP8	DEM	PU	44
M6.12	LMFDB full integration of algorithms, data and presentation (not so much verification)	WP6	R	PU	46
M1.7	Project Periodic Report (fourth year)	WP1	R	PU	48
M1.8	Project Final Report	WP1	R	PU	48
M3.5	Semantic-aware Sage interface to GAP.	WP3	O	PU	48
M3.8	Packaging for major Linux distributions	WP3	O	PU	48
M3.9	HPC enabled Sage distribution	WP3	O	PU	48
M7.2	OOMMF-NB evaluation report	WP7	R	PU	48

Milestones

The work in the OpenDreamKit project is structured by seven milestones, which coincide with the project meetings in summer and fall. Since the meetings are the main face-to-face interaction points in the project, it is suitable to schedule the milestones for these events, where they can be discussed in detail. We envision that this setup will give the project the vital coherence in spite of the broad mix of disciplinary backgrounds of the participants.

#	Name	WPs[1]/Deliverables involved	Mo	Means of Verif.
M1	**Initial Infrastructure**		1	Inspection
Set up the organizational infrastructure, in particular: Web Presence, project TRAC,. . .				
M2	**Consensus**		24	Inspection
Reach Consensus on the way the project goes				
M3	**Exploitation**		36	Inspection
The exploitation plan should be clear so that we can start on this in the last year.				
M4	**Final Results**		48	Inspection
all is done				

♠**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.* ♠

[1] The work package number is the first number in the deliverable number.

B.0.3.2 Work Package Descriptions

id=management
lead=PS

Work Package 1: Project Management				
Start: 0	**Activity Type:** MGT			
Site	*UPS*	USO	USTAN	**all**
Effort	48	2	48	**98**

Objectives

The objectives of this work package 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

This workpackage will perform all the activities related to monitoring of progress towards the project milestones shown on Page 33 and the deliverables listed on Page 30, 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.

♠**TO DO**: *MK: I would combine the first three into one “basic project infrastructure”*♠

Deliverables:
M1.1: (Month 1; nature: DEC, dissem.: PU) *Create tickets for all relevant tasks / deliverables*
M1.2: (Month 1; nature: DEC, dissem.: PU) *Internal and external mailing lists*
M1.3: (Month 1; nature: DEC, dissem.: PU) *Internal software repository*
M1.4: (Month 12; nature: R, dissem.: PU) *Project Periodic Report (first year)*
M1.5: (Month 24; nature: R, dissem.: PU) *Project Periodic Report (second year)*
M1.6: (Month 36; nature: R, dissem.: PU) *Project Periodic Report (third year)*
M1.7: (Month 48; nature: R, dissem.: PU) *Project Periodic Report (fourth year)*
M1.8: (Month 48; nature: R, dissem.: PU) *Project Final Report*

id=community lead=PS

Work Package 2: Community Building and Engagement				
Start: 5	**Activity Type:** ??			
Site	*UPS*	USHEF	USTAN	**all**
Effort	12	8	1	**21**

Objectives

The objective of this work package 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

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.

♠**TO DO**: *Neil: I have a series of Gaussian process summer schools and road shows that I'm organizing. These will also shift to more of a focus on data science across this year, I'd be happy to include these here if that's appropriate.*♠

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:
M2.1: (Month 6; nature: O, dissem.: PU) *Workshop 1*
M2.2: (Month 12; nature: R, dissem.: PU) *Report on community needs*
M2.3: (Month 18; nature: O, dissem.: PU) *Workshop 2*
M2.4: (Month 30; nature: O, dissem.: PU) *Workshop 3*
M2.5: (Month 42; nature: O, dissem.: PU) *Workshop 4*

id=component-architecture
read=CV

Work Package 3: Component Architecture					
Start: 3	**Activity Type:** ??				
Site	UPS	*UVSQ*	USHEF	USTAN	**all**
Effort	24	1	4	1	**30**

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

This package focuses on the structure of the components that make up a mathematical software and their interactions. Such components can be separate modules inside a unique software, or separate softwares interacting through library calls and/or through APIs (e.g.: web APIs). When combined together, they make up a full VRE.

The architecture of these software components must be:

- **Portable**, to support a wide range of platforms (mobile, desktops, cloud, . . .).
- **Modular**, so to ease installing, building, testing, and remixing.
- **Flexible**, so to adapt to different use cases: personal computation, HPC, parallel platforms, . . .
- **Open**, in the sense of *open source*, but also more importantly in the sense of clearly documented and open to the user who wants to understand its underpinnings and/or contribute to it. Indeed we must not forget that the working mathematicians needs to know what algorithms the software is going to run to solve a given problem.

id=portability

T1 **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 OS X), 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/GP`, `GAP`, `Singular`, `LinBox`) to this platform (?? ??.
- The deployment of a common infrastructure for multi-platform continuous integration (testing, building and distribution) will be addressed.

id=interface-systems

T2 **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.

We will build upon the work already done in the EU FP6 project 26133 "SCIEnce – Symbolic Computation Infrastructure for Europe" ♠**TO DO**: *fix and insert URL*♠ on the Symbolic Computation Software

Composability Protocol (SCSCP). SCSCP 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 is already supported by a number of computer algebra systems, including `GAP`, Macaulay2, Maple, TRIP and others. We will extend support for SCSCP to other relevant systems involved in OpenDreamKit. Through its API, 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`. 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 mapping `GAP`’s categories to `Sage`’s categories, we will:

- Implement a modular infrastructure for adapters, based on SCSCP, 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. id=mod-packaging

T3 **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 OpenDreamKit. 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 WP7 on 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. id=component-for-HPC

T4 **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 WP5.

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.

T5 **Improving the development workflow in mathematical software.** Truly open software must enable any actor to easily contribute his work to the community. Be it an experienced developer, or a student. Be it for a major software component, or for a piece of translation. All the systems involved in OpenDreamKit have developed their own workflows for contributing back, but those are almost exclusively geared toward experienced developers working on large components. When these workflows eventually reach their scalability limits, software development stagnates and major features are delayed. A well known example is given by Sage's TRAC server, where tickets can stay in "needs review" state for a long time before entering the code base. *Upstream* bug reporting and fixing is another major factor of slow development.

This task will seek new ways of accepting contributions to mathematical software in a scalable way. For example we will experiment with integrating bug reporting and contributing features right in the VRE (e.g., in SageMathCloud).

Deliverables:

M3.1: (Month 12; nature: DEM, dissem.: PU) *One-click install Sage distribution for Windows with Cygwin 32bits*

M3.2: (Month 24; nature: DEM, dissem.: PU) *One-click install Sage distribution for Windows with Cygwin 64bits*

M3.3: (Month 36; nature: P, dissem.: PP) *Continuous integration platform for multi-platform build/test.*

Sage's *buildbot* is a x86_64/Linux specific platform for continuous building and testing. We will investigate the possibility of evolving it towards a multi-platform tool, and opening it to other mathematical software.

M3.4: (Month 24; nature: O, dissem.: PU) *Support for the SCSCP interface protocol in all relevant components (Sage, GAP, etc.) distribution*

M3.5: (Month 48; nature: O, dissem.: PU) *Semantic-aware Sage interface to GAP.*

M3.6: (Month 12; nature: O, dissem.: PU) *Virtual images and containers* Creation and distribution of preconfigured cloud oriented virtual machines/containers (e.g. Docker images) for PARI/GP, Sage, ...

M3.7: (Month 24; nature: O, dissem.: PU) *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 ??

M3.8: (Month 48; nature: O, dissem.: PU) *Packaging for major Linux distributions* Make sure that Sage and all the components it depends on (including GAP, Linbox, PARI/GP, Singular, ...) have standard packages in the main Linux distributions.

M3.9: (Month 48; nature: O, dissem.: PU) *HPC enabled Sage distribution*

M3.10: (Month 24; nature: O, dissem.: PU) *Integration between SageMathCloud and Sage's TRAC server* id=UI

Work Package 4: User Interfaces									
Start: 24	**Activity Type:** ??								
Site	UPS	Logilab	UB	USHEF	USO	USTAN	JacU	UK	**all**
Effort	1	1	1	12	21	1	12	1	**50**

Objectives

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

Description

Project Jupyter is a set of open-source software projects for interactive and exploratory computing. These software projects help make scientific computing and data science reproducible and multi-language (Python, Julia, R, Haskell, etc.). The main application offered by Jupyter is the Jupyter notebook, a web-based interactive computing platform that allows users to create data- and code-driven narratives that combine live code, equations, narrative text, interactive dashboards and other rich media. These documents provide a complete record of a computation that can be shared with others.

♠Jupyter/... [**WRITE HERE**: Add references of Jupyter's use in Europe]♠ The Jupyter notebook is being used in all areas of academic (including University of California, Berkeley, Stanford, MIT, Harvard, Cambridge, Oxford, Imperial College, Southampton, Hamburg, Paderborn, Vienna, Paris, Katowice, Oslo♠All [**WRITE HERE**: Need more places using IPython in Europe here]♠) and government (NASA JPL, LBL, KBase, White House Hackathon) research as well as industry (Google, IBM, Facebook, Oracle, Otto Group, Microsoft, Bloomberg, JP Morgan, WhatsApp, O'Reilly, Quantopian, Logilab, GraphLab, Enthought, Continuum, Authorea, BuzzFeed, etc.) and journalism (538, New York Times, etc.). Because the architecture and building blocks of Jupyter are open, they are being used to build numerous other commercial and non-profit products and services. The Jupyter Notebook has between 500,000 and 1.5 million individual users worldwide.

In this work package, we will add new functionality to the Jupyter Notebook that fosters excellence in computational science and research. In particular, we will push towards reproducible and effective science by allowing to create structured documents (such as reports, books, thesis) from Notebooks, and by allowing those notebooks to be re-executed as self-contained regression tests. We will also unify the notebook infrastructure used in Sage with Jupyter, push forward dynamic documentation exploration capabilities, and work towards concurrent multi-user editing of Notebooks.

The last tasks in this work package are focused on an *application of the Jupyter notebook technology* for education and research.

♠Hans-Petter / Marcin? [**WRITE HERE**: Say something about notebook for education and planned demonstrators [notebooks]?]♠

To demonstrate the power of the Jupyther environment to to accelerate computational science, deliver better value for money and make computational science more robust, we will develop a state-of-the art Jupyter notebook interface and front end for a simulation package that is actively used in materials research by a wide range of scientists and engineers, in academia and industry. We have chosen the Object Oriented MicroMagnetic Framework (OOMMF) simulation package [1] as the target tool which is used to simulate magnetic nanostructures in over 1800 publications [2]. We use OOMMF-NB (for OOMMF NoteBook) as an identifier for this case study.

T1 **Uniform notebook interface for all interactive components** In this task, we will implement Jupyter interfaces for the interactive computation components of OpenDreamKit, including GAP, Pari, Sage, and Singular. A first release M4.1 will focus on basic functionality, and a second release M4.2 will cover advanced features like 3D graphics or transparent documentation browsing (as live worksheets whenever relevant). id=ipython-kernels

Sage itself will require a specific treatment as it already has a notebook interface. Its development started about at the same time as the Jupyter 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 Jupyter notebook was to take its time and build robust foundations from the ground up. The two projects have exchanged a lot, and the Jupyter 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 Jupyter notebook.
The Sage and Jupyter convergence M4.3 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.

id=notebook-collab

T2 **Notebook improvements for collaboration** In this task, we will further improve tools for collaboration between authors of shared Jupyter notebooks.
Version control tools, such as Git and Mercurial, have become an integral part of open and collaborative science and software. Version control tools allow reviewing proposed changes via diffing tools, and resolving conflicting changes with merge tools. Jupyter notebook documents, being text files, are relatively well suited to tracking in version control. However, being structured JSON data, diffing and merging are difficult. Tools shall be developed to provide better support for visual diffing and merging of Notebook documents, and integrated into existing version control workflows M4.5.
Given the interactive nature of Jupyter notebooks, live collaboration, where multiple authors work on the document simultaneously, as in Google Docs, is particularly desirable. The addition of potentially shared execution adds both value and challenge to live collaboration. Various projects have added some amount of live sessions from outside (SageMathCloud, Colaboratory), but outside the core project. There are many different aspects of collaboration to explore, including shared or separate execution for authors collaborating on a single notebook, UI to indicate not only authorship, but which author triggered which execution, and other challenges. Various avenues for live session collaboration will be explored for integration into Jupyter itself M4.7.

id=notebook-verification

T3 **Reproducible Notebooks** In this task, we will develop tools that allow to re-execute notebook documents with automated regression testing. The computed output will be compared against the stored output, and deviations reported as assertion errors.
Notebooks are used in a variety of contexts, like training and teaching material (tutorials, documentation, books) or computer experimentation logbooks, where reproducibility is critical: the notebooks shall remain functional and correct in the long run, even when the underlying computational software or infrastructure changes over time or across platforms.
This task is a critical step toward reproducibility, allowing the notebook author to get an immediate notice when, e.g., a backward incompatible change occurs. It becomes even possible to anticipate such situations upstream by including important notebooks directly in the automated test suite of the computational software, giving an easy way for casual users to contribute regression tests.
Technically speaking, Jupyter notebooks store outputs as rich mime-type structures, with JSON metadata. Using this metadata, it will be possible to express expectations of output, allowing more flexible and powerful tests than direct text comparison M4.4. Prior work has been done in Sage for ReST files, e.g. sage −t notebook.rst, and this model will be extended to notebooks.

id=dynamic-inspect

T4 **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]♠
- M4.8 (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 video on `http://www.larchenvironment.com/`) and sage-explorer.

  Ultimate goal: automatically generated LMFDB-style interfaces. Mention Knowls, as dynamic context-free items of knowledge

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

T5 **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 (T3).
- Collaborative editing.
- 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.*♠ id=oommf-python-interface

T6 **OOMMF case study: Create Python interface to OOMMF code**

First, we will identify best option for interfacing from Python to OOMMF core (C++) routines. The technical options include CTypes, Cython, Swig, and Boost-Python, all with particular advantages/disadvantages. Following analysis of the current OOMMF code layout and compilation model, we will use the most suitable tool, bearing in mind our ambition not to modify the OOMMF code so that the python interface we create remains functional and maintainable with minimal effort while the OOMMF core code is developed further by the OOMMF authors.
The interface will expose the raw C++ objects in Python, and for clarity we will refer to this interface as `OOMMF-py-raw`, to annotate that this gives access to OOMMF from Python but in a RAW way. Creation of this `OOMMF-py-raw` is technically doable as OOMMF had been written allowing to do this from Tcl. The `OOMMF-py-raw` library for Python provides access to the OOMMF functionality but requires some care when being used.
Secondly, we will create a user friendly Python library that combines the OOMMF-RAW capabilities we expect to become the main user interface to OOMMF in the medium term future. This will make use of object orientation to assist users in efficient and safe exploitation of the available facilities, following the design of the well-received high level Nmag simulation package [9] interface [22].
Once this is completed, several new features will be available to OOMMF users: (i) ability to drive OOMMF from Python, (ii) computational steering, and (iii) combination of OOMMF simulation with the existing Python eco-system of computational tools.
We illustrate the advantage of (iii) through an example: to solve the micromagnetic standard problem 3 [3], traditionally multiple OOMMF simulation runs would have to be conducted, and for each of those a new configuration file as to be written. Between these the size of the simulated geometry needs to be modified until two particular values of energy are the same. Given the new interface developed in this work package, this whole process can be replaced by one Python script that creates multiple OOMMF simulations, combined with a root finding method for the automatic iterative determination of the required simulation geometry.

Parallel in developing this, a set of unit tests is created that can be run periodically as regression tests. For all tasks relating to OOMMF-NB, documentation and tests are created simultaneously with the code. All codes, tests and documentation will be made available as open source.

We anticipate to start this task T6 in month 4, leading to deliverable M4.9.

id=oommf-py-ipython-attributes

T7 **OOMMF case study: Extend `OOMMF-py` with Jupyter notebook attributes and GUI templates** The web server based Notebook environment (Jupyter) allows to host, execute and document the Python-based OOMMF simulation in an executable document. In this interactive environment, the representation of objects can be overloaded, and can include representation of objects as text, as bitmaps or SVG files. We will create this functionality so that magnetisation vector field objects can be presented as a rendered 3d and 2d-view of the magnetisation field, and similar features for scalar fields such as field components and energies. This allows the interactive exploration and computational steering of the behavior of magnetic nanostructures. Depending on the development of 3d packages such as vispy, it may be possible to provide interactive data objects in the notebook.

Beyond that, the Jupyter Widgets allow the creation of graphical user interface (GUI) like elements, and we will generate code to display these widgets on demand to (i) set up micromagnetic simulation using a GUI, and (ii) assist in post-processing simulation results (deliverable M4.10). Not all OOMMF users are keen on using GUIs for simulation set up or post processing, but in particular new or infrequent users benefit significantly from this. Recent pilot work has shown that it is possible to make Jupyter Widgets interact with the python interpreter session and this allows to activate a GUI-like (widget based) interface when desired but to quickly return to the interpreter prompt, taking forward the results (data) from the GUI session [23] and providing a continuous path from scripting to GUI. We believe that having the ability to mix and match GUI-based and command driven analysis combines the best of both approaches and provides significant additional value.

id=oommf-tutorial-and-documentation

T8 **OOMMF-NB demonstrator: executable tutorial and documentation**

We will create documentation and a new tutorial on usage of OOMMF that introduces micromagnetic modeling in the new framework of OOMMF-NB, combined with complete documentation of the `OOMMF-Py` library. The documentation will be provided in form of executable Jupyter notebooks (deliverable M4.11).

The tutorial, in terms of content, will take guidance from the tutorial provided for Nmag [21] but tailored for the special simulation capabilities of OOMMF, and will introduce the special capabilities of the new IPython interface for OOMMF.

The output of this activity will deliver multiple benefits: providing a systematic introduction to `OOMMF-py` suitable for both those users (i) new to micromagnetic modelling and those (ii) new to the `OOMMF-py` interface. Because the documentation is developed in an Jupyter notebook, the documents are executable. For new learners this is a great simplification because they can skip through the given document and execute the given examples there and then: at the moment, this is a process of manually writing a script, or locating it in the directory structure of files, then executing this, subsequently opening and processing the data files, etc. In the new model, this end-to-end simulation will start from specifying the material parameters in the notebook (all of this is given in the tutorial), to running the simulation in the notebook to processing of computed data while the simulation runs (or subsequently) in the notebook; thus providing one virtual research environment, with all the associated benefits of making best use of the scientist's time using the tool and environment.

The documentation and tutorial will include a number of typical micromagnetic case studies that (i) demonstrate the correctness of the code by executing some of the micromagnetic standard problems and (ii) demonstrate the additional power gained by the IPython-based OOMMF interface. We expect this substantial, executable documentation to become the standard resource that introduces researchers to computational micromagnetics.

id=oommf-nb-ve

T9 **OOMMF-NB online and downloadable virtual environments** Recently, a TeMPorary Jupyter NoteBook has been made available (at http://tmpnb.org) that allows anybody to open this URL and use their very own Jupyter notebook for quick calculations and tests online. We will provide similar functionality but for a server that provides the OOMMF-NB software and OOMMF-NB documentation and tutorials so that the tutorial can be executed immediately on that web server; thus removing the barrier of having to install (the OOMMF and Jupyter notebook) code before being able to interactively drive and test a simulation system (deliverable M4.12).

We will further provide as open source the scripts as deliverable M4.13 that allow creation of virtual environments (such as vagrant scripts to generate VirtualBox [28] images, and Docker [7] containers). These virtual environments underpin the web hosted temporary OOMMF-NB service (we anticipate to use Docker on the web hosted service) but are also of use to those users who want to download a complete virtual machine (such as a virtualbox image) to run their simulations within that machine. The same virtual machine images can also be used for Cloudhosted computing services.

We request 3100 EUR (ex VAT) to purchase a machine to provide these services (shared memory, 16 cores, 64GB RAM, small solid state drive to make the system more responsive). This machine will also support the regression testing and continuous integration (see task WP8.T3). Setup and maintenance of the machine is part of this work task. id=cfd-vis

T10 **Visualization system for fluid dynamics data in web-notebook**

`This is in fact raw version - only my point of view (MK)`

The jupyter notebook is a very attractive environment for building user interfaces for research. On the other hand computational fluid dynamics (CFD) is one of the most demanding cases of scientific visualization. The amount of data can be tremendous, especially in 3d time- dependent problems computed in a distributed fashion over large scale computational clusters. On the other hand, interactive inspection of the simulation can be a valuable tool which accelerates research. In such an inspection one does not need to transfer and gather the full dataset at each timestep - it can give satisfactory feedback by getting computed fields on user request or to preprocess certain quantities like cross sections with some predefined frequency.

Implementing a specialized 3d visualization tool for viewing fluid simulations would be beneficial for researchers. It would allow to work purely inside web browser with access to a HPC cluster for CFD problem solving. This would in turn significantly lower the threshold for using distributed fluid dynamics code by eliminating tedious installation of software on own workstations. In this task we will implement a 3d live web notebook based visualization module for the Lattice Boltzmann solver which is under development at the University of Silesia: sailfish [http://sailfish.us.edu.pl/].

Sailfish is an advanced Lattice Boltzmann solver designed from the ground up for distributed system of GPU compute clusters. It is implemented predominantly in Python, it uses run-time code generation techniques to automatically build optimized code for CUDA and OpenCL devices. Since running sailfish requires specialized hardware, it is reasonable to use it on dedicated HPC installations. On the other hand, CFD requires interactive visualization and the Jupyter notebook turned out to be an excellent environment for running simulations and viewing results. The main drawback of this solution is lack of 3d visualization. While it is possible to work using cross-sections of a domain, it always comes to the point when dataset has to be downloaded to a workstation for analysis with 3d software like paraview.

Our target would be to move the whole typical fluid dynamics workflow to the notebook. The following steps will be implemented:

- a simple (regular Cartesian mesh based) geometry editor with options to impose and inspect boundary conditions; selecting the location of boundary conditions is currently a non-trivial process that involves direct manipulation of numpy arrays
- live monitoring of fluid flow in 3d which will communicate with running simulation
- the ex-post data analysis module

Deliverables:
M4.1: (Month 12; nature: O, dissem.: PU) *Basic Jupyter interface for GAP, Pari, Sage, Singular*
M4.2: (Month 12; nature: O, dissem.: PU) *Full featured Jupyter interface for GAP, Pari, Singular*
M4.3: (Month 12; nature: DEM, dissem.: PU) *Sage notebook / Jupyter notebook convergence*
M4.4: (Month 18; nature: O, dissem.: PU) *Facilities for running notebooks as verification tests*
M4.5: (Month 12; nature: O, dissem.: PU) *Tools for collaborating on notebooks via version-control*
M4.6: (Month 24; nature: O, dissem.: PU) *Jupyter extension for 3d CFD visualization*
M4.7: (Month 36; nature: O, dissem.: PU) *Exploratory support for live notebook collaboration*
M4.8: (Month 36; nature: DEM, dissem.: PU) *Exploratory support for semantic-aware interactive wid-*

gets providing views on objects represented and or in databases
communication with live computing process post simulation data analysis module visualization of vector and scalar fields editor for geometry and boundary conditions on regular meshes

M4.9: (Month 9; nature: O, dissem.: PU) *Python Interface to OOMMF*

M4.10: (Month 15; nature: DEM, dissem.: PU) *Jupyter notebook Interface for OOMMF (OOMMF-NB)*

M4.11: (Month 21; nature: DEC, dissem.: PU) *OOMMF-NB executable tutorial and documentation*

M4.12: (Month 24; nature: DEC, dissem.: PU) *OOMMF-NB dynamic web service available*

M4.13: (Month 24; nature: O, dissem.: PU) *OOMMF-NB virtual machine images available for download*

```
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.
```

id=hpc

Work Package 5: High Performance Computing							
Start: 36	**Activity Type:** ??						
Site	UPS	Logilab	UJF	UB	USTAN	UK	**all**
Effort	1	12	12	1	1	1	**28**

Objectives

The objective of this work package is to improve the performance of the computational components of OpenDreamKit, 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

As in all other areas of science, properly supporting massively parallel architecture is a major challenge. Many of the computational components in OpenDreamKit 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. The French ANR HPAC project (2012-2015) has also widely contributed to design parallel exact linear algebra kernels which is a core component for most HPC applications. The `LinBox` library, used in sage, has benefited from this experience on the multi-core processing aspects.

In this work package, we will build on this momentum to further implement HPC support in the components Tasks T1, T3, and T5.

♠**TO DO**: *transition*♠

Many of the computational components of OpenDreamKit 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 T7 and T8, we will also boost performance by further developing and applying such compilation tools, while keeping a high-level approach.

id=hpc-pari

T1 **PARI** `PARI` is a C library mainly oriented toward arithmetic and number theory.
It currently supports both POSIX threads or MPI but lacks interfaces for parallelism. More precisely, it should be easier from an external package or software (e.g. Sage) to better exploit `PARI` parallel features.
On the other hand, most basic algorithms in the PARI library (e.g. integer factorization) are currently implemented using only one core. To make better use of multi-core architecture and more generally parallel architectures, we will devise a generic parallelization machinery to allow individual implementations to scale gracefully between single core / multicore / massively parallel machines while avoiding code duplication.

id=hpc-gap

T2 **GAP** ♠SL/AK [**WRITE HERE**: Task around HPC/parallelism/perf in GAP]♠

id=hpc-linbox

♠**TO DO**: *deliverable*♠

T3 **Linbox** Most intensive mathematical computations rely heavily on exact linear kernels and their ability to harness parallel computers, grids or clusters. The `LinBox` library, already delivers high sequential efficiency for mathematical software such as `Sage`. The parallelization of the library for

multi-core architectures has been initiated in the A.N.R. HPAC project and successfully set the building blocks for high performance algebraic computing. The task here is to address the remaining challenges for the use of such kernels through a general audience mathematical software, such as `Sage`. A first aspect focuses on code design and domain specific languages allowing to expose an abstraction of the parallel infrastructure and the parallel features of the code through the stack of libraries, and support the composition of parallel routines. More generally the second aspect concerns the development of new parallel algorithms and implementations, that are still barriers in the development of High performance mathematical applications. Lastly, the third part addresses the specificities of distributed computing, with a close focus on communications and heterogeneous infrastructures.
♠**TO DO**: *deliverable*♠

T4 **Singular** The unique challenge of parallelizing Singular has been that it is a decades-old project, with a codebase exceeding 300,000 lines of code and an enormous existing investment of development effort. This makes a wholesale manual rewrite or reengineering approach infeasible.
We therefore use a multi-pronged approach: First, we have created automated source-to-source translation tools that take existing C/C++ code as input and generate thread safe code as output. Secondly we are also adding facilities to the C/C++ code and the Singular interpreter to safely access shared memory. These facilities ensure in particular that common pitfalls of concurrent programming, such as data races and deadlocks, cannot occur. For this, we leverage approaches that have already been successfully used for HPC-GAP and whose principles are well-understood.
To supplement the above existing work, we propose to add very fine-grained parallelism to some key components of Singular. These include writing a multi-threaded implementation of the Singular multivariate polynomial arithmetic, of the main quadratic sieve implementation for integer factoring and parallelisation of the FFT based integer and dense polynomial multiplication algorithm. These key components are used extensively for Singular's overall workload, including in the Groebner basis engine and polynomial subsystems. Performance increases through fine-grained parallelisation of key components such as these will provide extensive benefits to virtually all users of Singular on multi-core machines. id=hpc-singular

T5 **MPIR** MPIR (Multiple Precision Integers and Rationals) is the core library in Sage for bignum arithmetic. It is used extensively by a majority of the core C/C++ libraries in Sage, and by Sage directly via Cython. MPIR is a fork of the GMP (Gnu Multi-precision) library, with many independent implementations of the core algorithms (including a faster FFT and division code, better superoptimisation on some common 64 bit processors and native MSVC support). It consists of around 250,000 lines of code, much of which is assembly primitives and very low level, highly optimised C code.
Maintenance of MPIR is not merely a matter of updating the build system. Rather, every time a new processor is released by AMD, Intel, Sparc or ARM, significant development has to be invested in hand-optimising and then superoptimising assembly code for the new processors. This gives up to a 12x performance increase over optimised C code, due to the specialised nature of bignum arithmetic, which is in some sense a worst case for C compilers. Indeed without continuous effort, MPIR would not even run on new operating systems and processors, let alone run fast. This is a unique problem that assembly libraries have.
As a successful and key component of Sage, we believe it is time to invest in maintenance and improvement of MPIR by hiring an assembly expert who can work full time on the project after MPIR's lead assembly expert sadly passed away recently. Significant challenges exist, such as optimising for SIMD instruction sets. Without investment into maintenance, assembly superoptimisation, processor support, fat binary support, etc. this key component of Sage will fall behind, to the detriment of Sage as a whole and the numerous other standalone libraries that make use of MPIR. id=hpc-combi

T6 **HPC infrastructure for combinatorics** ♠FH [**WRITE HERE**: Task around HPC infrastructure for combinatorics]♠
Several members of the projects are experts in combinatorics a field where Sage is clearly a world leader. This particular research field has several specific features which makes it interesting from the HPC point of view:

- The main obstacle is combinatorial explosion which stop many experiment very early. Algorithm of exponential complexity are extremely common.

- Embarrassingly parallel problem are extremely common.
- Problem that are not embarrassingly parallel very often reduce to a large tree exploration.

To help the researcher, it is crucial to minimize the extra work needed to get a parallel program from a serial one in these simple situations. Through this task we will provide a practical and concrete as well as highly demanding use case for the infrastructure developed. In particular, we aim to test on real problems the benefit of tasks T7, ??, ??, and ??.

In a second and more exploratory direction, some experiments shows that the large tree exploration problem is very easily solved in C++ using the new Intel `Cilk++` technology. We would like to explore the possibility to interface smoothly `Pythran`, `Cython` and `Cilk++`.

id=pythran-cython

T7 **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 called from `Cython` when relevant ?? 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 ??

This work will be achieved through a close collaboration between the `Pythran` developers hired for OpenDreamKit 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.

id=pythran-sage

T8 `Pythran` **for** `Sage` **and** `Sage` **Users** o

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 he added in `Sage` to improve user kernels using the `Pythran` compiler ??

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 required a continuous integration of `Pythran` to ensure its capabilities ??

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 T7, 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 ??

T9 **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:

M5.1: (Month 12; nature: R, dissem.: PU) *Library design and domain specific language exposing LinBox parallel features to $\mathtt{Sage}$*

M5.2: (Month 24; nature: DEM, dissem.: PU) *Algorithms and implementations. Library maintenance and close integration in mathematical software*

M5.3: (Month 36; nature: DEM, dissem.: PU) *Implementations of algorithms using distributed memory et heterogenous architectures: clusters and accelerators*

M5.4: (Month 3; nature: DEM, dissem.: PU) *Parallelise the relation sieving component of the Quadratic Sieve*

M5.5: (Month 9; nature: DEM, dissem.: PU) *Implement a parallel version of Block-Wiederman linear algebra over GF2 and the triple large prime variant.*

M5.6: (Month 12; nature: DEM, dissem.: PU) *Take advantage of multiple cores in the matrix Fourier Algorithm component of the FFT for integer and polynomial arithmetic,and include assembly primitives for SIMD processor instructions (AVX, Knight's Bridge, etc.), especially in the FFT butterflies*

M5.7: (Month 12; nature: DEM, dissem.: PU) *Parallelise the Singular sparse polynomial multiplication algorithms*

M5.8: (Month 12; nature: DEM, dissem.: PU) *Parallel versions of the Singular sparse polynomial division and GCD algorithms.*

M5.9: (Month 12; nature: DEM, dissem.: PU) *Extend the existing assembly superoptimiser for AVX and upcoming Intel processor extensions*

M5.10: (Month 24; nature: DEM, dissem.: PU) *Ongoing support of Intel, AMD, ARM, Sparc processors and new Operating System versions*

M5.11: (Month 12; nature: DEM, dissem.: PU) *Provide FAT binary support for all new x86_64 processors, build system maintenance and improvements to tuning, profiling and testing subsystems*

M5.12: (Month 3; nature: DEM, dissem.: PU) *Turn the Python prototypes for tree exploration into production code, integrate to Sage.*

M5.13: (Month 12; nature: DEM, dissem.: PU) *Explore the possibility to interface smoothly Pythran, Cython and Cilk++.*

M5.14: (Month 24; nature: DEM, dissem.: PU) *Refactor the existing combinatorics Sage code using the new developed Pythran and Cython features.*

M5.15: (Month 6; nature: DEM, dissem.: PU) *Implement Pythran runtime support in Cython when they are implemented instead of using default implementation.*

M5.16: (Month 3; nature: DEM, dissem.: PU) *Improve Pythran runtime support to automatically take advantage of multi-cores and SIMD instruction units.*

M5.17: (Month 2; nature: DEM, dissem.: PU) *Facility to compile Pythran compliant user kernels.*

M5.18: (Month 1; nature: R, dissem.: PU) *Ensure interoperability of Pythran with Python and its packages.*

M5.19: (Month 12; nature: DEM, dissem.: PU) *Make Pythran typing better to improve error information.*

idaedksbases

Work Package 6: Data/Knowledge/Software-Bases					
Start: 1	**Activity Type:** ??				
Site	USHEF	UW	*JacU*	UZH	**all**
Effort	12	3	36	1	**52**

Objectives

The objectives of this work package are: to design and implement interfaces that can be used for a wide range of mathematical data and to standardise metadata allowing for interoperability, searching, documentation, versioning and visualisation.

Description

Mathematics is the only science that has not yet benefitted greatly 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 with respect to [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 reveal this richness. This weakness prevents the formulation of new conjectures, the testing of new hypotheses, and generally an exploratory approach to mathematical data. The past has shown that such an approach can 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 Millenium Problems;
- the Monstrous Moonshine conjecture finds its origin in a numerical coïncidence between dimensions of representations of the Monster group and coefficients of the j-function, and its conclusion eventually led to Borcherds’ Fields medal.

Therefore to facilitate future advances, we need ways to represent DKS in the same systems, and – since current computational/experimental mathematics involve extensive DKS – we need a new kind of ”database”, which we will call 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 individual mathematicians or a distributed collaboration to manage and use such interlinked 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, to serve as exemplars to other projects) and in return will substantially enhance their capabilities. Prerequisites should be kept to a minimum (depending on contributors' and users' needs and goals), and in particular would not require any background in databases to contribute new data or perform queries.

T1 **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. id=data-assessment

T2 **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 OpenDreamKit, drawing on existing emerging standards. id=data-design
We will organise a workshop associated to this task.

T3 **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. id=data-triform

T4 **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: id=data-foundationCAS

- 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

T5 **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. id=data-OEIS

T6 **FindStat Case Study (triformal theories)** In this task we would develop triformal theories for the FindStat project to test the design from T4. Similarly to the previous task, in this case study, we first develop a thorough OMDoc/MMT model, which should only involve a handful of MMT theories (combinatorial collections, maps, statistics,...), each with a few hundred realisations. Together with ♠POD [**WRITE HERE**: WP4]♠, this will again allow for easier knowledge management services, and in particular improved search services. id=data-findstat
This Task will be co-developed with T4, it will validate the design of triformal theories and be iterated to test the design changes.

id=data-LMFDB

T7 LMFDB **Case study (triformal theories)** In this task we would develop triformal theories for an exemplary part of the LMFDB project to test the design from T4. We will identify a fragment of the LMFDB 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 T4. (decrease importance of verification as opposed to interoperability)

id=data-memo

T8 **Memoization and production of new data** Many CAS users run large and intensive computations, for which they want to collect the results while simultaneously working on software improvements. Sage currently has a limited cached_method, that is not persistent across sessions and does not enable to publish the result or share it with a smaller group of collaborators. We propose to use, extend and contribute back to some established persistent memoization infrastructure, such as `python-joblib`, `redis-simple-cache` or `dogpile.cache`. The caching should apply recursively to lower level functions, and should be trivial to setup and configure for the end user: in a single line, the user only needs to select an existing function and maybe provide some additional semantic information, and has the option to change the defaults for a few parameters, such as the backend (shared dropbox folder, remote directory, database, git repository, ...). The interface could be through a Python decorator. Additionally, it should be easy to launch a data-bot to populate the database, all the versioning and provenance tracking should be handled (user, algorithm, software version, ...), and the system should have useful data properties (atomicity, merging, and error detection).

Deliverables:

M6.1: (Month 12; nature: DEC, dissem.: PU) *Conversion of existing and new databases to unified interoperable system*

- Polytopes in Polymake
- graphs, graph properties
- Finite groups (Max)
- Lattices

M6.2: (Month 24; nature: O, dissem.: PU) *Shared persistent memoization library for Python/Sage* Recomputation?, Ease of publishing, importing, ...

M6.3: (Month 9; nature: R, dissem.: PU) *Workshop Report*

M6.4: (Month 12; nature: R, dissem.: PU) *Design of Triform (DKS) Theories (Specification/RNC Schema/Examples)*

M6.5: (Month 24; nature: O, dissem.: PU) *Implementation of Triform Theories in the MMT API.*

M6.6: (Month 12; nature: DEC, dissem.: PU) *Python/Sage Syntax Foundation Module in OMDoc/MMT*

M6.7: (Month 24; nature: O, dissem.: PU) *Python/Sage Computational Foundation Module in OMDoc/MMT*

M6.8: (Month 36; nature: O, dissem.: PU) *Python/Sage Declarative Semantics in OMDoc/MMT*

M6.9: (Month 12; nature: R, dissem.: PU) *LMFDB deep modelling: Fragment Identification & Initial Model Design*

M6.10: (Month 18; nature: R, dissem.: PU) *LMFDB Data vs. Knowledge vs. Software Validation*

M6.11: (Month 36; nature: O, dissem.: PU) *LMFDB Algorithm verification wrt. a Triformal theory*

M6.12: (Month 46; nature: R, dissem.: PU) *LMFDB full integration of algorithms, data and presentation (not so much verification)*

M6.13: (Month 9; nature: O, dissem.: PU) *Heuristic Parser for the OEIS*

M6.14: (Month 18; nature: R, dissem.: PU) *Cross-Validation for OEIS DKS-Theories*

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

id=social-aspects
lead=UO

Work Package 7: Social Aspects				
Start: 1	**Activity Type:** ??			
Site	*UO*	USHEF	USO	**all**
Effort	1	8	5	**14**

Objectives

♠**TO DO**: *Neil: Can we add something about scientific repeatability and sharing of mathematical models? I'm very keen to see systems where I can share a mathematical abstraction with a scientist in such a user-friendly way that they feel compelled to interact with it and develop their understanding of the modeling more. This might be an appropriate place to put this idea.*♠

The primary objective of this work package is to analyse the "mutual crowdsourcing" phenomenon occurring in the framework of development and maintenance of an open-source virtual research environment. We will focus on identifying the incentives for all participants of the system, that would encourage sustained development of the most important parts. To this end, we will use ideas from the recent research on crowdsourcing in algorithmic mechanism design. While doing so, we will apply outcomes to a case study system. Further, we will study development models for an academic free software ecosystem, and analyse how they facilitate the mathematical process behind relevant algorithms and databases. Finally, we will look into the questions of trust in results produced by computations done by VREs, and ways to mitigate the issues arising there.

Description

Crowdsourcing is a recent phenomenon in software development in particular, but not only there. E.g. crowdsourcing of mathematical ideas occurs in online mathematics communities, such as Mathoverflow [19] and in Polymath Projects [**polymath**]. "Mutual crowdsourcing" is the main driving force in developing and maintaining of any large-scale open-source virtual research environment.

Open source projects tend to be fragile, in the community sense, and suffer from disagreements that ultimately result in forks and the resulting repetition of effort. We will analyse this in a setup of cooperative game theory, and try to design a finely tuned systems of incentives and rewards for contribution, to increase the stability of the community and its useful output.

We will focus on identifying the appropriate incentives for all participants of the system, that would encourage sustained development of the most important parts of the system. To this end, we will use ideas from the burgeoning field of mechanism design [20] and in particular on recent research on crowdsourcing in algorithmic mechanism design [26]. While doing so, we will apply outcomes to a case study system—Sagemath. We will apply preference and opinion aggregation techniques [5] to develop a community prioritisation scheme for Sagemath bugs and features requests, which are presently being maintained on the Sagemath TRAC server [4] and implement them as a TRAC [27] add-on.

As well, we will study various development models for an academic free software ecosystem, and analyse how they facilitate the mathematical process behind algorithms being designed and implemented, and databases of experimental data and test cases being created and expanded. Trusting results of computer computations is crucial for usability; channels for communicating bug reports and fixes need to be carefully analysed from social point of view. Commercial closed source computer algebra and other computational systems often fail to react to bug reports in a timely manner, and seemingly are falling into the short-sighted trap of hiding bugs from potential and current users [8], Open source systems are only marginally better in this way, as recent computer security scares, such as the one around Bash [6], indicate. A game-theoretic analysis of this situation will be attempted.

T1 **Modern Distribution of Scientific Output** ♠**TO DO**: *Neil: a suggested task in this domain, it's an area I've been thinking about a lot and it could fit in here.* ♠ The current model for distribution of scientific output stems from an era when the printing press was dominant. The process has become formalized through peer review and publication of journals. The PDF format for distribution of documents reflects the status quo, that a scientific paper is a written as if for printing and remains an unchanging document. In scientific blogging we are seeing that more rapid propagation of ideas can occur when the constrains of the printed format are released, however, there is a lack of formalization that means attribution of ideas and commentary run amiss. We will develop tools and ideas for distribution of scientific knowledge that don't rely on a static format and allow for the full spectrum of scientific debate. The tools will encourage proper credit attribution through encouraging sharing of attribution for ideas, software and data. ♠NL [**WRITE HERE**: One deliverable simple idea would be to enable the construction of 'Living Posters' which can be interacted with, targeted at large touch screens.]♠

T2 **Survey and collection of needed data** We will survey the data needed to assess development models of large-scale academic open-source projects, such that the probable correlation between the size of the atomic contribution vs. the speed of the contribution making it into the code, and collect appropriate statistical data. The latter will require non-trivial amount of programming work, even only for the test system, Sagemath. id=datacollection

T3 **Collective decision making in development** Currently development of open-source academic software is task-driven, where tasks (also known as tickets) are posted on a website, and their priorities are set in an ad hoc manner. Whereas the latter might be good enough for simple bug fixing, for more elaborate task this often leads to delays etc. We would like to investigate an voting-driven approach, where the priorities are being voted on by the developer community, and possibly the people who completed tasks are incentivised in some form (e.g. by "karma points", as on Mathoverflow). id=decisionmaking

T4 **OOMMF case study: Evaluation** We will start to make the OOMMF-NB code (see WP4.T6 to WP4.T8) available as open source to micromagnetic community research groups as early as possible (WP8.T3) to gather feedback on its usefulness over a period as long as achievable. We will use this case study data to support our study of social aspects of mutual crowdsourcing.
A survey will be developed and used to gather user input and feedback on usefulness of the provided capabilities, with particular focus on where OOMMF-NB enables new and better science, where it allows to do work more effectively and reproducibly, and the role of community contributions. We will further gather suggestions for improvements, and this feedback will be evaluated continuously and where possible re-acted on. We will `target OOMMF-NB` users that attend workshops (see task WP8.T4). All results and insights will be summarised in a report (M7.2) to share the lessons learned from this user interface application example, and be made publicly available. Where possible, we will report our findings at relevant workshops/conferences.

Deliverables:
M7.1: (Month 12; nature: ??, dissem.: PU) *...*
M7.2: (Month 48; nature: R, dissem.: PU) *OOMMF-NB evaluation report*

id=dissem

Work Package 8: Dissemination					
Start: 18	**Activity Type:** ??				
Site	USHEF	USO	USTAN	US	**all**
Effort	8	10	1	24	**43**

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 OpenDreamKit project;
- defining individual exploitation plans; and,
- managing existing and new intellectual property.

Description

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

T1 **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 WP3 Component Architecture.

T2 **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 (??, 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 (?? ??, and the project will organise at least one open technical workshop each year.

♠**TO DO**: *Neil: This coming year I will do four workshops, although the number may be smaller by the time we are up and running. An additional component that could be integrated here is a 'JMLR Monthly Notices' section that I recently gained approval for. This is a new section the leading machine learning journal that we'll aim to focus on typically smaller advances (less than a full paper). A little like a 'Nature Notices' section but much more informal and open. We don't yet know the best way of doing it, but the mechanism might be similar to ways we'd want to disseminate DreamKit results.*♠

id=dissemination-of-oommf-nb-virtual-environment

T3 **Demonstrator: Open source dissemination of OOMMF-NB virtual environment** OOMMF-NB (see WP4.T6 to WP4.T8) is build on top of OpenDreamKit and demonstrates the power of the environment in a production environment of active OOMMF users. The source code of OOMMF-NB will be made available as open source on public repository hosting sites (such as github/bitbucket), and announced to the community via appropriate mailing list (mumag, magpar, nmag, mumax, micromagnum) and other means. We will encourage participation of the micromagnetic community in the maintenance and development of the tools, and allow time to train users to join the OOMMF-NB project to drive towards a self-sustained OOMMF-NB project at the end of the funding period.

To underpin this process, we will set up a publicly accessible Jenkins/Travis continuous integration (CI) system to (i) run regression tests (from WP4.T6 and WP4.T7) routinely when the OOMMF-NB code changes or OOMMF core code changes and (ii) re-execute notebooks (from WP4.T8) and use as regression test (making using of the outcome of task WP4.T3). This will test user-contributions automatically, and can automatically create virtual environments for download that containing the latest versions (M8.5).

We will create a manuscript for journal publication (M8.6), summarising the approach taken and experience gathered so far. An important point of this publication is to provide a reference that can be cited by publications making use of the new OOMMF-NB, allowing tracking of uptake of this technology in the medium and long run.

id=dissemination-of-oommf-nb-workshops

T4 **OOMMF-NB open source dissemination workshops**

In this task, we run a series of workshops (M8.8) at major international meetings to disseminate the open source OOMMF-NB tool (see T3) in the micromagnetic community.

We will prepare the software stack for the workshops, develop teaching materials, and do the planning (location, advertising, registration) and delivery of workshops at the 4 most significant international meetings taking place in the appropriate time frame[2] each of which tends to attract around 1500 attendees. Depending on demand, multiple workshops will be given per conference, i.e. one workshop per evening. We anticipate to travel with two people (Hans Fangohr and PDRA or other support staff) to be able to teach effectively and deliver hands-on training as part of the workshop. The taught material will include (i) use of the OOMMF-NB workflow model to support effective and reproducible computational science. To support to make the OOMMF-NB project self-sustaining and to maximise the value of this initial develop, we will also teach an (ii) introduction to the standard techniques for contribution to open source software (version control, pull requests, testing frameworks). In addition, all teaching materials, including videos, will be made available on a website.

We request 500 EUR room hire per meeting to support delivery of these workshops directly at the international meetings.

id=ibook

T5 **Demonstrator: interactive books**

One of important elements of VRE is a common flexible format which enables the creation of large structured documents. There are many known solutions to that problem, but they usually compromise the interactivity of the notebook interface and quality of desktop publishing software like LaTeX.

Recently, few approaches tried to bring both interactivity and the typographic features. Modestly tagged markup language implementation DocOnce targets the problem of reusability of the document source code. The MathBook XML is a lightweight XML application for authors of scientific articles, textbooks and monographs extensively using Sage single cells for interactive elements. Sphinx documentation software has been successfully applied for creation of interactive books containing Sage cells using sagecell plugin.

The technical aspects of format for interactive publications is a subject of the task “Structured documents” in WP4.T5. In this task we will demonstrate usability of the results of WP4.T5 in creation of scientific textbooks. Three interactive books will be created:

- Nonlinear Processes in Biology
- Classical Mechanics
- Problems in Physics with Sage/python

The choice of those particular topics has been made for the sake of maximal diversity. The “Nonlinear Processes in Biology” will heavily use numerical ODE and PDE integration with addition to the classical approach. Classical Mechanics will demonstrate power of CAS systems working “on par” with

[2] Anticipated major meetings are 61st Conference on Magnetism and Magnetic Materials October 31-November 4, 2016, New Orleans, Louisiana; 62nd Conference on Magnetism and Magnetic Materials November 6-10, 2017, Pittsburgh, Pennsylvania; 21st International Conference on Magnetism (ICM 2018) July 16–20, 2018, San Francisco, California; 14th Joint MMM-Intermag Conference January 14-18, 2019, Washington, DC). Each of those meetings is one week long, and serves as a focal point of networking for the european and international research community.

numerical tools. The last example will focus mostly on collaborative editing and modularity of content which is produced using VRE technologies.
The main research aspect for this task will be to find a way for efficient application of computational tools in problems solving and analysis. In addition the work-flow will be optimized in order to explore the full potential of VRE environment.
In particular we will answer following questions:

- How to create a monograph at reusing independently working building block of text and code?
- When the interactive worksheet should be used and when executable code cell inside interactive textbook is sufficient?
- What are best tools and practices for using single source for with many output targets?
- How to collaboratively write reusable course materials?
- How students can benefit from using VRE as working tool?

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”[3], where Python has been recently selected as the language to learn programming[4].

Deliverables:
M8.1: (Month 36; nature: DEM, dissem.: PU) *Demonstrator: Nonlinear Processes in Biology interactive book*
M8.2: (Month 40; nature: DEM, dissem.: PU) *Demo: Classical Mechanics interactive book*
M8.3: (Month 12; nature: DEM, dissem.: PU) *Demo: Problems in Physics with Sage v1*
M8.4: (Month 30; nature: DEM, dissem.: PU) *Demo: Problems in Physics with Sage v2*
M8.5: (Month 32; nature: DEC, dissem.: PU) *OOMMF-NB public repository and continuous integration*
M8.6: (Month 36; nature: R, dissem.: PU) *OOMMF-NB publication manuscript completed*
M8.7: (Month 44; nature: DEM, dissem.: PU) *Demo: Problems in Physics with Sage v3*
M8.8: (Month 42; nature: O, dissem.: PU) *OOMMF-NB workshops delivered*

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

[3] http://en.wikipedia.org/wiki/Classe_pr\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{e\global\mathchardef\accent@spacefactor\spacefactor}\accent19e\egroup\spacefactor\accent@spacefactorparatoire_aux_grandes_\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{e\global\mathchardef\accent@spacefactor\spacefactor}\accent19e\egroup\spacefactor\accent@spacefactorcoles

[4] See the “Annexe” at http://www.education.gouv.fr/pid25535/bulletin_officiel.html?cid_bo=71586

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

B.0.3.3 Management Structure and Procedures

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

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

Explain why the organizational 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).♠

Risk	Level	Mitigation measures
Recruiting qualified personnel	Medium	whenever possible, we coordinate with currently running projects to rehire personnel with well known high level competences
slmhnlnhfnhs	hsfhs	ghshsh

♠**TO DO**: *An open source architecture allows risk sharing: collaborating in the early stages of research could help an early detection, and, by consequence, reducing risks.*♠

♠**TO DO**: *But: since Open Source softwares are freely accessible, security and privacy issues are a concern. Anytime a resource is shared, there is greater risk of unauthorised access and contaminated data. Providers must demonstrate security solutions, which should include physical security controlling access to the facility and protection of user data from corruption and cyber attacks.*♠

B.0.3.4 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*

♠

♠DP/SL/AK/... [**WRITE HERE**: Proofread consortium as a whole]♠
The consortium brings together:

1. Lead or core developers of some of the major open source computational components for pure mathematics and applications: `GAP` (Saint-Andrews, Oxford), `LinBox` (Grenoble), `MPIR` (Kaiserslautern), `PARI` (Bordeaux, Versailles), `Sage` (Orsay, Versailles, Bordeaux, Oxford, Zürich), `Singular` (Kaiserslautern).
2. Leaders of a major mathematics online database: `LMFDB` (Warwick, Zürich).
3. Experts in mathematical knowledge management (Bremen).
4. The lead developers of the Virtual Research Environment for mathematics `SageMathCloud` (Seattle).
5. Experts and major "promoters" of the `Jupyter` collaborative user interfaces for interactive and exploratory computing (Southhampton, Simula, Sheffield) in a variety of scientific domains.
6. Lead authors of `Jupyter` based interactive teaching resources for physics.
7. Lead developers of Pythran and experts in numerical code optimization/parallelization (Logilab, Grenoble)
8. A company specialized in open-source based Database and Scientific Computing for the industry (Logilab); it develops in particular its own virtual environment `Simulagora`.
9. ♠DP [**WRITE HERE**: Social aspect in Oxford]♠.

The first role of 1, 2, and 3 will be to design the OpenDreamKit VRE architecture adapted to the needs in pure mathematics and applications (Objective 1), as defined in collaboration with 9. They will be supported in this objective by 4 and 5 which bring respectively expertise in the key technologies `SageMathCloud` and `Jupyter`. They will also benefit from the **diversity of real world use cases from all areas of scientific computing, in academia and industry** provided by 5, 8, and 6.

All the aforementioned participants will be themselves primary users of the OpenDreamKit VRE, produce demonstrators thereof (Objective **??**), and participate actively to the dissemination (Objective 9).

5 will host and mentor core `Jupyter` developers to improve this key technology (Objective **??**), while 1, 2, and 3 will update the mathematical software components (Objective 4). 7 will be a key asset for the latter, providing expertise in massive parallelism and HPC, and bringing in and further developing the specific `Pythran` optimization technology.

8 will further bring in expertise in semantic databases, distribution of large software, and open source based business models.

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

Finally 1, 4, and 5 also have strong experience in community building and engagement (Objective 3).

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**: *Linbox: recruitment of previous ANR developer*♠

B.0.3.5 Resources to be Committed

♠NT [**WRITE HERE**: Rework this section]♠

♠**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**: *Table 3.4.a: insert here table from Figure 3, and transpose; see Table 3.4.a in the word template*♠ ♠**TO DO**: *The work package leader will usually have the largest investment*♠

Table 3: Overview: Resources to be committed (all in €)

♠**EC Commentary**: *Please complete the table below for each participant if the sum of the costs for' travel', 'equipment', and 'goods and services' exceeds 15budget 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.♠

B.0.4 Members of the Consortium

B.0.4.1 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.*

♠

B.0.4.2 USTAN: UNIVERSITY OF ST ANDREWS (UK)

The Centre for Interdisciplinary Research in Computational Algebra (CIRCA) fosters research at the interface of Mathematics and Computer Science including abstract and algorithmic algebra and combinatorics, formal languages and automata, mathematical software and constraint programming. Our success is founded on the close integration of theoretical and algorithmic research and the development and use of state-of-the-art software.

In 1997, CIRCA became the centre of the development of GAP after the retirement of under Prof. J. Neubüser who initiated the system in mid-1980s in Aachen. This move was supported by EPSRC, EU and Leverhulme grants. Nowadays, as one of the centres of GAP development, CIRCA has excellent contacts with developers and users worldwide.

CIRCA's output includes first class results in Pure Mathematics and in Computer Science, recognised by our highly-cited publications in top international venues in both disciplines and our widely used research software. Beyond the individual international connections of the investigators and research staff, CIRCA as a centre has national importance and international standing. CIRCA has been selected to host major conferences such as CP 2010, BCC 2009, PP 2007, and the "Groups St Andrews" series in 2005 and 2013.

Curriculum vitae

Steve Linton is a Professor of Computer Science at St Andrews. He has worked in computational algebra since 1986 and has helped coordinate the development of GAP since its move from Aachen in 1997. He personally wrote key features of GAP, such as workspaces and exception handling, and has overseen the development and releases of the whole system. He directed CIRCA from 2000–2013. He is an editor of AAECC[5]. He has been PI of four major EPSRC grants and coordinated the EU project SCIEnce. He is the general chair of ISSAC 2015, the main conference in computer algebra.

Alexander Konovalov is a Senior Research Fellow in CIRCA and has worked on GAP for more than 10 years. After holding the fellowship at the Vrije Universiteit Brussel in 2006, researching computational group ring theory, he moved to St Andrews in 2007 to join EU project SCIEnce. He leads many aspects of the GAP project, including release preparation, regression testing and liaison with package authors. He has authored 38 papers and 8 GAP packages, and co-organised a number of events, most recently the LMS/EPSRC Short Instructional Course in Computational Group Theory, the HPC-GAP workshop (2013), and the Summer School on Experimental Methodology in Computational Science Research (2014). He is an editor of Journal of Software for Algebra and Geometry and a Fellow of the Software Sustainability Institute.

Publications, products, achievements

1. S. Linton, K. Hammond, A. Konovalov, C. Brown, P.W. Trinder, H.-W. Loidl, P. Horn and D. Roozemond, Easy Composition of Symbolic Computation Software using SCSCP: A New Lingua Franca for Symbolic Computation. J. Symbolic Computation, 49 (2013), 95–119.

Previous projects or activities

1. **Multidisciplinary Critical Mass in Computational Algebra and Applications** EP/C523229 2005–2010.
2. **SCIEnce: Symbolic Computation Infrastructure for Europe** (FP6 eRII3-CT-026133) 2006–2011.
3. **HPCGAP: High Performance Computational Algebra and Discrete Mathematics** EP/G055181 2009–2014, four sites.

Significant infrastructure

CIRCA provides hosting for the GAP website and ftp-servers, runs the automated system to check, fetch and test updates of GAP packages, uses the Jenkins continuous integration system to run regression tests for the development version and release candidates on its computer infrastructure, and manages GAP releases preparation.

[5] Applicable Algebra and Error Correcting Codes

B.0.4.3 UPS: UNIVERSITÉ PARIS SUD (FR)

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, Inria), 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 (PI) 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. type=PI PM=6

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, paralellism, . . .), specialized research libraries.

Viviane Pons (leadPI) 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 including a talk at PyCon US 2015. Before committing herself to research, she spent two years working in industry as a Java and web developer. type=leadPI PM=6

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 (PI) 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. type=PI PM=12

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 (incarnated as `MuPAD-Combinat` until 2008); 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[6]. Along the way, he coauthored part of the proposal for NSF `Sage-Combinat` grant OCI-1147247.

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.*♠

[6] http://sagemath.org/library-publications-combinat.html, http://sagemath.org/library-publications-mupad.html

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.

B.0.4.4 UK: UNIVERSITY OF KAISERSLAUTERN (DE)

B.0.4.4.1 University of Kaiserslautern

type=PI

Prof. Dr. Wolfram Decker (PI) 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.

Dr. William Hart William Hart is a postdoctoral researcher at the University of Kaiserslautern. He is the lead developer of the Flint and MPIR projects as well as the main author and maintainer of the BSDNT bignum library, the Nemo and ANTIC libraries and a contributor to various other software packages.

Before coming to Kaiserslautern, held a prestigious five year Career Acceleration Fellowship "Algorithms in Algebraic Number Theory" at Warwick University in the UK. He has been involved in a number of high performance computing records, including computation of congruent numbers (subject to the BSD conjecture) up to a trillion (10^{12}).

William is the main author of the FFT code for multiplication of large integers and polynomials in both MPIR and Flint, which are used extensively by the Sage, Singular and Macaulay 2 computer algebra systems.

The main focus of William's research has been in algorithms for fast arithmetic,fast integer and polynomial factorisation and to algebraic number theory, including computation of modular equations and class invariants.

Publications, products, achievements

1. SINGULAR computer Algebra system.
2. Wolfram Decker is coordinator of the DFG Priority Project SPP1489 *Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory'.*
3. FLINT AND MPIR C libraries for number theory and bignum arithmetic.
4. William Hart held an EPSRC Career Acceleration Fellowship EP/G004870/1 from 2008-2013, *Algorithms in Algebraic Number Theory*

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.

B.0.4.5 UO: UNIVERSITY OF OXFORD (UK)

The University of Oxford is in the top ten universities worldwide in the Shanghai 2013 and 2014 rankings (the Universities of Cambridge and Oxford are the only non-US university there). It employs over 10,000 staff and has a student population of over 21,000. Most staff are directly appointed and managed by one of the University's 130 departments or other units within a highly devolved operational structure - this includes 5,900 'academic related' staff (postgraduate research, computing, senior library, and administrative staff) and 2,820 'support' staff (including clerical, library, technical, and manual staff). There are also over 1,600 academic staff (professors, readers, lecturers), whose appointments are in the main overseen by a combination of broader divisional and local faculty board/departmental structures.

The Department of Computer Science is one of the longest established Computer Science departments in the country. Formerly known as the University Computing Laboratory, it is home to a community of world-class research and teaching. Research activities encompass core Computer Science, as well as Security, Algorithms, computational biology, quantum computing, computational linguistics, information systems, software verification and software engineering.

The department is home to undergraduates, full-time and part-time Master's students, and has a strong doctoral programme. Students are highly skilled and motivated, and as practice shows, easily start contributing to open-source projects such as Sagemath.

Curriculum vitae

PM=2

Ursula Martin Professor Ursula Martin has recently joined the University of Oxford, where she holds a Professorship, in conjunction with a Senior EPSRC Fellowship, on a joint arrangement between the Department of Computer Science and the Mathematical Institute. Her current research concerns social and computational techniques for creating mathematics, building on a significant track record at the interface of mathematics and computing. Prior to this she worked at Queen Mary University of London, where as Vice Principal for Science and Engineering she led strategic change, and was active in knowledge transfer activities and developing young staff.

Her work is characterized by strongly interdisciplinary collaboration in new problem domains at the interface of mathematics and computer science, identifying novel interactions between theory and practice, with real-world problems inspiring scientific advance. Major achievements include results linking randomness and symmetry, new unifying explanations of the power of computational logic, and new practical techniques for using computational logic and algebra in industry.

The work to be undertaken in the Work Package 5 (Social Aspects) fits very well into her current project, which concerns crowdsourced mathematics: the overarching goal is to understand and extend the human and computer creation of mathematics. It is mostly funded by her 2014 EPSRC Advanced Fellowship (EPSRC awards only one or two of these annually in Computer Science) is a partnership of industry, government and international academia.

PM=2

Edith Elkind is an Assoiciate Professor at the Department of Computer Science of the University of Oxford. She is a leading expert in algorithmic game theory, computational social choice and voting theory, and in multiagent systems. In 2014 she received an ERC Starting Grant for the project “Algorithms for Complex Collective Decisions on Structured Domains”. The settings considered in the present proposal provide a natural application domain for tools to be developed in the ERC project.

Edith holds a PhD (2005) from Princeton; before coming to Oxford in 2013 she held a Singapore National Research Foundation Fellowship totalling € 1.6M, graduating 3 PhD students and supervising a number of postdoctoral researchers. She chaired/is chairing program committees of major international conferences, such as AAMAS, and is an associate editor of a number of major journals, such as Artificial Intelligence Journal.

PM=24

Dmitrii Pasechnik is a Senior Research Fellow at the Department of Computer Science of the University of Oxford, where he also holds a Lectureship at Pembroke College. Before moving to Oxford in 2013, he taught mathematics for 8 years in Nanyang Technological University (Singapore). While there, he was successful in receiving individual grant funding totalling over € 500K, graduated 2 PhD students, supervised post-doctoral researchers, and co-organized a 2-months research program at Singapore Institute for Mathematical Sciences on a range of topics in computational mathematics, involving over 100 participants.

He works on a wide area of interconnected topics, related to computational algebra and optimization, combinatorics, algorithm, symbolic computing, and game theory, and authored over 70 papers on these topics, several of them using Sage and/or its components, such as GAP.

He is an active Sage developer, and regularly contributes, himself or together with his undergraduate or graduate students, new or improved Sage interfaces to various mathematical packages and databases. He taught a Sage-based undegraduate course while at Nanyang, and has good understanding of the overall Sage development process, as well as of development of other open-source software and databases, including their social/community aspects.

Publications, products, achievements

1. U. Martin. Computational logic and the social. J.Logic & Computation (2014) [doi:10.1093/logcom/exu036]
2. U. Martin, A. Pease. Mathematical Practice, Crowdsourcing, and Social Machines, in Springer LNCS vol. 7961 (2013) [doi:10.1007/978-3-642-39320-4_7]
3. G. Chalkiadakis, E. Elkind, M. Wooldridge. Computational Aspects of Cooperative Game Theory, Morgan & Claypool, 2011 [doi:10.2200/S00355ED1V01Y201107AIM016]
4. E. Elkind, D. Pasechnik, Y. Zick. Dynamic weighted voting games, in Proc. AAMAS 2013 [http://dl.acm.org/citation.cfm?id=2484920.2485
5. S.H. Chan, H.D.L. Hollmann, D. Pasechnik. Sandpile groups of generalized de Bruijn and Kautz graphs and circulant matrices over finite fields, J. Algebra 421(2015) [doi:10.1016/j.jalgebra.2014.08.029]

Previous projects or activities

1. U. Martin holds an EPSRC Senior Fellowship, 2014-2017, to study the social machine of mathematics.
2. E. Elkind will hold an ERC Starter Grant, 2015-2020, to study and develop algorithms for collective decision making on structured domains.
3. D. Pasechnik supervised students contributing, and significantly contributes himself, to Sagemath, to OEIS, and to GAP.

Significant infrastructure

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

B.0.4.6 UB: Université Bordeaux (FR)

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

Curriculum vitae

PM=12,salary=2500 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

PM=12

Karim Belabas Professor of Mathematics in Bordeaux (France) since 2005, Karim Belabas is a senior researcher in Number Theory, with particular interest in computational and effective aspects. Karim has published about 25 articles in international journals, including papers in Duke and Compositio (one of which co-authored with Manjul Bhargava, 2014 Fields Medal), and edited a book on "Explicit methods in number theory".

Karim was head of the Pure Math teaching department in Bordeaux from 2010 to 2014 and is vice-head of the Institut de Mathématiques de Bordeaux since 2015. He has held a grant from the French ANR worth 200 kEuros (ALGOL project, 2007–2011) and was part of a 2.5MEuros Marie-Curie Research Training Network (GTEM project 2006–2010); he was responsible for three deliverables and supervision of an early stage researcher during her PhD in the Work Package "Constructive Galois Theory". He has (co-)organized 8 international conferences, including a special Trimester at IHP in 2004 and an influential recurrent workshop on "Explicit methods" in Oberwolfach (every two years since 2007) and 5 PARI/GP Ateliers. He has (co-)supervised 11 PhD students and about 15 masters students.

Karim is a leading computational number theorist in France. He is the project leader for the PARI/GP free computer algebra system since 1995, which has had a major impact on hundreds of publications. He is one of the system's main developers (about 60000 lines of code written, most of the documentation, and 1300 bug-tracking tickets authored).

PM=6

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

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: Plafrim and Avakas.

B.0.4.7 UJF: UNIVERSITÉ JOSEPH FOURIER (FR)

The Université de Grenoble partner gathers two teams form the Laboratoire Jean Kuntzmann (CASYS team, with Jean-Guillaume Dumas and Laurent Fousse) and the Laboratoire d'Infor- matique de Grenoble (MOAIS, with Thierry Gautier, Clément Pernet and Jean-Louis Roch). The CASYS team is specialized in Algebraic computations, cryptology, codes and hybrid symbolic- numeric dynamical systems. The MOAIS team is specialized in programming and scheduling design on distributed resources for applications based on interactive simulation. The software developed by this partner is significant. It includes K AAPI , L IN B OX and also G IVARO (C++ library for arithmetic and algebraic computations), PAC (Parallel Algebraic Computations), Galet (Matrix multiplication schedule generator), FFSpMV (sparse matrix-vector product over finite fields), PaloAlto (cryptology on curves and cryptanalysis toolbox), Taktuk (tool for de- ployment of parallel remote executions of commands to a large set of remote nodes), etc.

Curriculum vitae

PM=6

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 Quantfier elimination, Optimization, Linear Algebra and Polynomial Systems and he is the director of the French ANR program on High-Performance Algebraic Computations.

PM=12

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 developper 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.

PM=12

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)]♠

B.0.4.8 Logilab: LOGILAB (FR)

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 [**WRITE HERE**: Expand the end of Logilab's site description]♠

Curriculum vitae

♠Logilab [**WRITE HERE**: Add Logilab CV's]♠

Publications, products, achievements

1. CubicWeb (mention of prize at DataConnexion#3)
2. publi Brainomics

Previous projects or activities

1. ASWAD (eu)
2. KIDDANET (eu)
3. PYPY (eu)
4. OpenHPC (fr/FUI)
5. BRAINOMICS (fr/ANR)
6. Mention Debian development experience

Significant infrastructure

B.0.4.9 UW: University of Warwick (UK)

The Mathematics Institute at the University of Warwick was ranked 23rd worldwide in the 2013 QS world university subject rankings, and third in the UK in the 2014 Research Excellence assessment. 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

PM=3

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 LMFDB project.

Cremona has been a 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 `mwrank` 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.

B.0.4.10 UWS: University of Washington at Seattle (US)

♠WS [**WRITE HERE**: Expand description of University of Washington]♠

The University of Washington is a large research university, which receives over a billion dollars a year in federal funding.

Curriculum vitae

type=R
PM=0

William Stein (R) William Stein is a number theorist at the University of Washington who does research on elliptic curves and modular forms. He founded the Sage project in 2004, wrote a large amount of the code of Sage, organized dozens of Sage Days workshops, and won the *2013 Richard Dimick Jenks Memorial Prize for Excellence in Software Engineering applied to Computer Algebra* for his work on Sage. He has written dozens of research papers and also published two mathematics textbooks, and has two more books under contract to publish within the next year.

Stein's main current software development project is SageMathCloud, which is an open source web application he launched in 2013 that lets people use Sage and much more collaboratively online. In addition to growing the user base and implementing new functionality, he is also creating a company to make SageMathCloud better supported and more widely available and scalable.

Publications, products, achievements

Previous projects or activities

1. William Stein has received many grants from the National Science Foundation for work on `Sage` and number theory.
2. There have been an average of 4 Sage Days workshops at UW every year since 2006.

Significant infrastructure

University of Washington hosts a farm of 30 computers (well over 500 cores) dedicated to the `Sage` project.

B.0.4.11 JacU: Jacobs University Bremen (DE)

Jacobs University is a private Anglo-Saxon style research university. It opened in 2001 and has an international student body (1320 students from 115 nations as of 2011). The KWARC (KnoWledge Adaptation and Reasoning for Content [13]) Group headed by *Prof. Dr. Michael Kohlhase* specializes in knowledge management for STEM. Formal logic, natural language semantics, and semantic web technology provide the foundations for the research of the group.

Curriculum vitae

type=leadPI
PM=6

Michael Kohlhase (leadPI) Dr. Michael Kohlhase is full professor for Computer Science at Jacobs University Bremen and an associate adjunct professor at Carnegie Mellon University.

He studied pure mathematics at the Universities of Tübingen and Bonn (1983 - 1989) and continued with computer science, in particular higher-order unification and automated theorem proving (Ph.D. 1994, Saarland University).

His current research interests include knowledge representation for mathematics, inference-based techniques for natural language processing, and computer-supported education. He has pursued these interests during extended visits to Carnegie Mellon University, SRI International, and the Universities of Amsterdam, Edinburgh, and Auckland.

Michael Kohlhase is recipient of the dissertation award of the Association of German Artificial Intelligence Institutes (AKI; 1995) and of a Heisenberg stipend of the German Research Council (DFG 2000-2003). He was a member of the Special Research Action 378 (Resource-Adaptive Cognitive Processes), leading projects on both automated theorem proving and computational linguistics. Michael Kohlhase was trustee of the Conference on Automated Deduction (CADE), Mathematical Knowledge Management (MKM), and the CALCULEMUS conference, he is a member of the W3C Math working group, president of the OpenMath Society, and the general secretary of the Conference on Intelligence Computer Mathematics (CICM).

PM=6
type=PI

Florian Rabe (PI) is a post-doctoral research fellow at Jacobs University Bremen. He completed his PhD in 2008 and his habilitation in 2014 and holds the venia legendi.

He has worked on the formal representation and management of mathematical knowledge for 10 years. He was a lead researcher in the LATIN project (2009-2012), which produced a highly modular and integrated library of formal languages for knowledge representation. He is currently a principal investigator in the OAF project, which builds on LATIN to produce an archive of libraries of formal mathematical knowledge.

He is the creator and main developer of the MMT language and system, which are the backbone of both LATIN and OAF. MMT has been developed for 8 years with contributions from > 10 people and currently consists of $> 30,000$ lines of Scala code.

He served in the organization committee of 2 and the program committee of 6 international conferences (2 as track chair) on intelligent computer mathematics, and has organized 4 international workshops on module systems and libraries for mathematical knowledge. He has authored 65 research papers (11 in international journals) and has supervised 17 undergraduate and graduate theses.

PM=12
type=Res

Christian Maeder (Res) Dr. Christian Maeder is a research developer at Jacobs University. He has extensive experience in designing and implementing logic-based software system. He is the lead implementor of the HETS system.

type=JRes
PM=24

Mihnea Iancu (JRes) Mihnea Iancu is a third-year doctoral student at the KWARC group. He is the lead implementor of the MathHub.info system. He has worked extensively on the representation for formal and informal mathematical knowledge in the MMT syste.m

Relevant previous experience: The KWARC group is the lead implementor of the OMDoc (Open Mathematical Document) format for representing mathematical knowledge [14] and redeveloped its formal core in the OMDoc/MMT format [25]. The latter has been implemented in the MMT system [24, 25] which provides efficient implementations of the computational primitives such as type checking, flattening, and presentation at a logic/foundation-independent level. The group has developed services powered by such semantically rich representations, different paths to obtaining them, as well as platforms that integrate both aspects. *Services* include the adaptive context-sensitive presentation framework provided by the MMT API and the semantic search engine MathWebSearch[16, 15].

Semantic services can be integrated into the documents generated from OMDoc/MMT representations, making them into "active documents", i.e. documents that are interactive and adaptive to the user and situation. For *obtaining* rich content, the group investigates assisted manual editing [12] as well as automatic annotation using linguistic techniques [10]. Finally, KWARC has developed the MathHub.info portal a community-based library and knowledge management system for flexiformal libraries, which can be used for semantic publishing and eLearning [17, 18, 11].

The OMDoc/MMT knowledge representation format and the MathHub.info system will an important basis for the developments Work Packages 4 and 6.

Michael Kohlhase has initiated and led the CALCULEMUS! IHP-Research and Training Network and participated in the FP6 IST MoWGLI (Mathematics on the Web: Get it by Logic and Interfaces) project, the FP6 CSA Once-CS (Open Network of Centres of Excellence in Complex Systems), The FP7 EDC project WebALT (Web Advanced Learning Technologies).

Specific expertise:

- Modeling formal structures of mathematical knowledge in a web-scalable way
- Transforming large collections of legacy scientific publications to semantically structured markup
- Designing user interfaces for authoring and interacting with mathematical knowledge

B.0.4.12 UZH: UNIVERSITÄT ZÜRICH (CH)

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

type=leadPI
PM=6

Paul-Olivier Dehaye (leadPI) 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 formulas 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.

B.0.4.13 Simula: SIMULA RESEARCH LABORATORY (NO)

Dedicated to tackling scientific challenges with long-term impact and of genuine importance to real life, Simula Research Laboratory (Simula) offers an environment that emphasizes and promotes basic research. At the same time, we are deeply involved in research education and application-driven innovation and commercialisation.

Simula was established as a non-profit, limited company in 2001, and is fully owned by the Norwegian Ministry of Education and Research. Its research is funded through competitive grants from national funding agencies and the EC, research contracts with industry, and a basic allowance from the state. Simula?s operations are conducted in a seamless integration with the two subsidiaries Simula School of Research and Innovation and Simula Innovation.

At its outset, the laboratory was given the mandate of becoming an internationally leading research institution within select fields in information and communications technology. These fields are (i) communication systems, including cyber-security; (ii) scientific computing, aiming at fast and reliable solutions of mathematical models in biomedicine, geoscience, and renewable energy; and (iii) software engineering, focusing on testing and verification of mission-critical software systems, and on planning and cost estimation of large software development projects. Recent evaluations state that Simula has met its challenge and is an acknowledged contributor to top-level research in its focus areas. Specifically, in the 2012 national evaluation of ICT research organized by the Research Council of Norway and conducted by an international expert panel, Simula received the highest average score (4.67) on a 1-5 scale among all evaluated institutions. In comparison, the national average was 3.38. Only five of the 62 research groups evaluated were awarded the top grade (5), and two of these five groups are located at Simula.

Simula is currently hosting one Norwegian Centre of Excellence, Center for Biomedical Computing (2007-2017), and one Norwegian Centre for Research-based Innovation, Certus (2011-2018). In addition, we participate as research partner in another Centre for Research-based Innovation, Centre for Cardiological Innovation (2011-2018), hosted by Oslo University Hospital. These two center-oriented schemes are the most prestigious funding instruments offered by the Research Council of Norway.

PM=24

Curriculum vitae

Hans Petter Langtangen is director of Center for Biomedical Computing at Simula Research Laboratory, a Norwegian Center of Excellence doing inter-disciplinary research in the intersection of mathematics, physics, computer science, geoscience and medicine. Langtangen is on 80% leave from a position as professor at the Department of Informatics, University of Oslo.

Langtangen received his PhD from the Department of Mathematics, University of Oslo, in 1989, and then worked at SINTEF before being hired as assistant professor at the University of Oslo in 1991. After being promoted to full professor of mechanics at the Department of Mathematics in 1998, he moved in 1999 to a professorship in computer science. In the period 1999-2002 he also held an adjunct professor position at the Department of Scientific Computing at Uppsala University in Sweden. The Simula Research Laboratory was formed in 2001, and Langtangen has since then worked with research and management at this laboratory. The scientific computing activity at Simula has been awarded the highest grade, Excellent, by five panels of top-ranked international scientists in the period 2001-2012.

Langtangen's research is inter-disciplinary and involves continuum mechanical modeling, applied mathematics, stochastic uncertainty quantification, and scientific computing, with applications to biomedicine and geoscience in particular. He has also been occupied with developing and distributing scientific software to make the research results more widely accessible and help accelerate research elsewhere. For over three decades he has been very active with teaching and supervision.

The scientific production consists of 4 authored books, 3 edited books, about 60 papers in international journals, about 60 peer-reviewed book chapters and conference papers, and over 130 scientific presentations. The publications cover fluid flow, elasticity, wave propagation, heat transfer, finite element methods, uncertainty quantification, and implementation techniques for scientific software. Langtangen is on the editorial board of 7 journals and serves as Editor-in-Chief of the leading SIAM Journal on Scientific Computing. He is also a member of the Norwegian Academy of Science and Letters.

Publications, products, achievements

1. A. Logg, K.-A. Mardal, G. N. Wells et al. Automated Solution of Differential Equations by the Finite Element Method, Springer (2012). [doi:10.1007/978-3-642-23099-8]
2. P. E. Farrell, D. A. Ham, S. W. Funke, and M. E. Rognes. Automated Derivation of the Adjoint of High-Level Transient Finite Element Programs. SIAM J. Sci. Comput. 35-4 (2013), pp. C369-C393
3. H.P. Langtangen. A Primer on Scientific Programming with Python. Texts in Computational Science and Engineering, Springer (2014), 792 pp..

Previous projects or activities

1. The Centre for Biomedical Computing, a Norwegian Centre of Excellence, awarded by the Research Council of Norway. Duration: 2007-2017. Budget: 75 MNOK (10 MEUR).
2. The FEniCS Project (www.fenicsproject.org) Duration: 2007–on-going.

Significant infrastructure

The fully owned Simula subsidiary Simula Innovation handles pre-commercial innovation projects, creation and follow-up of company spin-offs, and general support for entrepreneurs.

B.0.4.14 USO: UNIVERSITY OF SOUTHAMPTON (UK)

The University of Southampton (UoS) is one of the leading universities in the United Kingdom, was founded in 1952 and is a member of prestigious Russell Group of UK Universities. UoS has more than 19,000 undergraduate students and 4,000 postgraduates and is an excellent venue for conducting cutting-edge research and for providing high quality education. The university is truly international, drawing students from over 130 different countries and benefiting from a wide and varied culture. It is ranked in the top 1% of universities worldwide (QS world university rankings 2014-15) and in the top 15 of research led universities in the UK, and is participating in a high number of collaborative research projects and related initiatives. UoS has a successful track record of industrial collaborations and is at the centre of a cluster of local high technology companies. It has an enviable track record in the generation of patentable work, with a portfolio of over 350 patents. To ensure the impact of its research projects, University of Southampton's Research & Innovation Services (R&IS) is responsible for professional protection of IP and supporting commercial development with industry. R&IS has had considerable success, licensing annual revenue in excess of € 1million and launching twelve successful spin-out companies since 2000. UoS has a strong track record of working in European projects, especially within the Framework Programme. The EC 6th FP7 Monitoring Report ranked UoS 17th out of all higher and secondary education organisations for number of FP7 participations during 2007-2012. Throughout the FP7 UoS has received € 132M in research grants and has been involved in 319 projects, including 63 ICT and 8 INFRASTRUCTURES Collaborative Projects. In 2013/14 alone UoS has received over € 181.5M in research grants and contracts, including over € 16.3M from the European Commission.

The Faculty of Engineering and the Environment (FEE) is one of the lead engineering faculties in Europe, educating a range of professionals and generating research of the highest quality. Southampton's world-leading engineering ranking is confirmed by being ranked first in the UK for the volume and quality of the research in Electronic Engineering, Electrical Engineering and General Engineering in the latest Research Excellence Framework (REF) 2014.

FEE brings together a wide range of disciplines, offering undergraduate and postgraduate programmes in audiology and environmental science as well as acoustical, civil and environmental, mechanical, and aeronautical/astronautical engineering and ship science. It consists of 370 research postgraduate students and 340 academic and research staff. FEE also hosts the University Technology Centres and Research Framework Agreements with key partners including: Airbus, Rolls-Royce, Lloyd's Register, Microsoft and Network Rail. FEE has a strong background in working on international research projects, including 84 EU FP7 projects worth over € 28M. In 2013/14 only FEE has received about € 50M in research grants and contracts, of which over € 1.7M from EU funding programmes.

Curriculum vitae

PM=6
type=PI

Hans Fangohr (PI) Hans Fangohr is Professor of Computational Modelling at the University of Southampton. He has studied Physics with specialisation in Computer Science and Applied Mathematics, gained his PhD in High Performance Computing (2002) in computer science and has since worked on the development of computational tools and application of those in interdisciplinary projects in science and engineering.

He heads the University's interdisciplinary Computational Modelling Group (`http://cmg.soton.ac.uk`), and has more than 100 publications on development of computational methods and applied computer simulation in magnetism, superconductivity, biochemistry, astrophysics and aircraft design.

In 2013, he has attracted € 5m from the UK's Engineering and Physical Sciences Research Council (EPSRC) together with additional moneys from industry and his University of Southampton to fund the € 12m Centre for Doctoral Training in Next Generation Computational Modelling (ngcm.soton.ac.uk) in the UK. This flagship activity will train about 75 PhD students (10 to 20 starting every year, first cohort started in September 2014) in the state-of-the-art and best-practice in computational modelling, the programming of existing and emerging parallel hardware and to apply these skills and tools to carry out PhD research projects across a range of topics from Science and Engineering. The centre has chosen IPython as a key tool to deliver this teaching, document and communicate computational exploration and drive reproducible computation to push for excellent computational science.

Hans Fangohr has led the development of the Open Source Nmag software (http://nmag.soton.ac.uk), which provides a finite-element micromagnetic simulation suite to a community of material scientists, engineers and physicists who research magnetic nanostructures in academia and industry. He has designed the package in 2005 so that it has an IPython-compatible Python interface, to make the workflow of using the simulation package as accessible as possible to scientists without substantial computational background. He has extensive experience in micromagnetic simulation tool development and use, and due to this an outstanding understanding of the requirements for computational workflows in this micromagnetic research community.

He has deep interest in excellence and innovation in learning and teaching. He has been awarded the prestigious Vice Chancellor's teaching award ($£1000$) three times (in 2006, 2010, 2013) for initiating and realising three separate innovations in the university's teaching delivery of computational engineering, and has been voted "best lecturer" and "funniest lecturer" of the year by the students. Other Universities in the UK and elsewhere have adopted his teaching methods and materials. He has attracted grants to further develop learning and teaching activities, and given invited talks at international meetings on efficient learning and teaching of computational methods.

Hans Fangohr is chairing the UK's national Scientific Advisory Committee for High Performance Computing.

NN (R) We will hire a post-doctoral senior research fellow to carry out the work at Southampton, under the leadership of and together with Hans Fangohr. The fellow will have a background in computational science, ideally in micromagnetics, combined with solid IPython and `Jupyter` Notebook experience, and past experience of software engineering. We further require good communication and team working skills, and in particular interest and skill in the development of education materials to best support this part of the project. type=R PM=32

Publications, products, achievements

1. Open Source micromagnetic simulation framework Nmag, http://nmag.soton.ac.uk, Thomas Fischbacher, Matteo Franchin, Giuliano Bordignon, Hans Fangohr: *A Systematic Approach to Multiphysics Extensions of Finite-Element-Based Micromagnetic Simulations: Nmag IEEE Transactions on Magnetics **43**, 6, 2896-2898 (2007)*
2. Other open source contributions to the micromagnetic simulation community: OVF2VTK, higher order anisotropy extensions to OOMMF, OVF2MFM, summarised at http://www.southampton.ac.uk/ fangohr/software/index.html
3. H. Fangohr. *A Comparison of C, Matlab and Python as Teaching Languages in Engineering* Lecture Notes on Computational Science **3039**, 1210-1217 (2004)

Previous projects or activities

1. EPSRC Doctoral Training Centre in Complex Systems Simulations (http://icss.soton.ac.uk), jointly funded by EPSRC and the University of Southampton, € 14m, (2009–2018)

Significant infrastructure

1. The University of Southampton hosts the largest university owned Supercomputer "Iridis 4" in the UK (12,300 cores, 250 TFlops), the hardware (€ 3.75m) is refreshed every 3 years.
2. A community of 200 academics and over 500 researchers and doctoral students are users of this facility and provide a wide network pushing forward excellent computational science in the context of solving real world problems.
3. EPSRC Centre for Doctoral Training in Computational Modelling in the United Kingdom, € 12m. (http://ngcm.soton.ac.uk), (2013–2022)

B.0.4.15 UVSQ: UNIVERSITÉ DE VERSAILLES SAINT-QUENTIN (FR)

Université de Versailles – Saint-Quentin-en-Yvelines

PRiSM Laboratory. The research teams of the PRiSM laboratory (Parallélisme, Réseaux, Systèmes et Modélisation) are involved in two main scientific themes of UVSQ: Mathematics and Computer science on one hand, "Design, Modelization and Implementation of Systems" on the other hand. These two directions are not separated from each other, as shown by many collaborations with other labs, and the participation of many PRiSM teams to both directions. Within the "Mathematics and Computer Science" theme, the PRiSM teams study cryptology and security, models for algorithms and operational research. All the teams also participate to the "Design, Modelization and Implementation of Systems" theme, with a particular focus on communication systems (networks and telecommunication), embedded systems, mobile systems, high speed networks, and database systems.

PRiSM is home to the "Cryptology and Information Security". In its research activities, the cryptography team aims at widely covering the various themes of academic research in cryptology, public key and secret key cryptography, cryptanalysis, security of implementations, number theory, multivariate cryptography, hash functions, etc. The cryptology team brings its specificity in the computer science courses at UVSQ and, since several years, the university offers several teaching programs with a part devoted to cryptology and information security. In particular, the research graduate program "Applied Algebra" offers a full course in cryptology. It has been complemented by a professional graduate program, called SeCReTS (Security of Contents, Networks, Telecommunications and Systems). Many activities of the team, require the use of advanced computer algebra. For this, the team has a long history of using computer algebra systems (GAP, Pari, Maple, Magma, ...). In recent years, with the arrival of young researchers, and with the affirmation of Sage in research and teaching, the team has moved from a pure user perspective to a contributor one, taking active part in the development of computer algebra software.

Curriculum vitae of the investigators

Luca De Feo got his PhD in 2010 at Ecole Polytechnique. He was appointed Maître de Conférences at Versailles-St-Quentin-en-Yvelines University in 2011. His research interests cover Algorithmic Number Theory, Computer Algebra, Cryptology and Automated deduction, and he has already published 8 papers in international journals or refereed international conferences.

He is an active Sage contributor, with a dozen of tickets co-authored and about as much reviewed. He is also active in promoting the use of Sage for research and for teaching: most of his papers feature a publicly available Sage implementation, he teaches Sage to undergraduate and graduate students, he participates and organizes various events for the introduction of Sage to beginners and young researchers.

Nicolas Gama got his PhD in 2009 at École Normale Supérieure. He was appointed Maître de Conférences at University of Versailles-St-Quentin-en-Yvelines in 2010. His research interests cover Lattice reduction algorithms, Theory of computer sciences, Algorithmic Number Theory, and Cryptology. He has already published 12 papers in international journals or refereed international conferences.

He developed a fork of the NTL library to ease the development of parallel lattice algorithms, and added various blockwise lattice primitives, tools like high dimensional gaussian sampling over lattices and modulo lattices which can be directly used to implement the most recent lattice-based schemes. This fork is scheduled to be merged with the main branch of NTL, and the wrapper library for Sage should then be updated accordingly.

Publications, achievements

Recent publications:

1. A. Becker, N. Gama and A. Joux; Solving shortest and closest vector problems: The decomposition approach. ANTS 2014.
2. L. De Feo, J. Doliskani, É. Schost; Fast arithmetic for the algebraic closure of finite fields. ISSAC '14. ACM, 2014. pp 122-129.
3. N. El Mrabet and N. Gama, Efficient Multiplication over Extension Fields, WAIFI 2012.
4. L. De Feo, É. Schost; Transalpyne: a language for automatic transposition. ACM SIGSAM Bulletin, 2010, 44 (1/2), pp. 59-71.

Software:

1. newNTL. It is a high-performance, portable C++ library providing data structures and algorithms for manipulating signed, arbitrary length integers, and for vectors, matrices, and polynomials over the integers and over finite fields. http://www.prism.uvsq.fr/~gama/newntl.html.
2. FAAST, a C++ library for Fast Arithmetic in Artin-Schreier Towers. http://github.com/defeo/FAAST.

Previous projects or activities

Current grants:

1. ANR CLE (2013-2017): Cryptography from Learning with Errors. The goal is to propose fast and secure symmetric protocols based on the LWE problem.
2. DIGITEO project ARGC (2013-2016): “Fast arithmetic for geometry and cryptology”. The project explores fast algorithms and implementations for algebraic geometry and curve-based cryptography.
3. DIGITEO project IdealCodes (2014-2016): IdealCodes (http://idealcodes.github.io/) spans the three research areas of algebraic coding theory, cryptography, and computer algebra, by investigating the problem of lattice reduction.

B.0.4.16 US: UNIVERSITY OF SILESIA (PL)

The University of Silesia in Katowice was established in 1968. Now, with 12 faculties and several interdisciplinary schools and centers, over 30000 students and over 2000 academic staff the University is one of the largest in Poland. Students are educated at three educational levels: Bachelor, Master and Doctoral and their achievement are accumulated using European Credit Transfer and Accumulation System (ECTS). Located in the heart of Upper Silesia, Poland's old industrial region with distinct history and cultural identity, the university attracts many scientists and students.

The origins of the *Faculty of Mathematics, Fhysics and Chemistry* date back to the academic year 1968/1969 and coincide with the establishment of the University of Silesia. One of the largest university units, the faculty incorporates, as its name indicates, three separate departments: mathematics, physics and chemistry, each with several divisions and subdivisions carrying out the research and educational activities. There are over 1900 students, both full-time and part-time, educated at three educational levels: Bachelor's, Master's and Doctoral. The Faculty is entitled to grant doctoral degrees in the natural sciences. The Faculty staff consists of 243 academics who are both teachers and researchers.

Curriculum vitae

PM=12

Marcin Kostur is an assistant Professor at the Institute of Physics, author of over 40 publication cited over 1000 times in the field of statistical physics, solid state physics (Josephson Junction dynamics), microfluidics and biophysics. He is experienced in application of GPU architecture to numerical simulations of stochastic processed in physics. His recent computiational interests are focused at the Open Source project "Sailfish" - HPC implementation of Lattice Boltzmann Method on GPU.

He is leader of two projects incorporating computations to the science education:

- Computing in high school science education - iCSE4schools, project funded by Erasmus+, Key Action 2 - Strategic Partnerships", (budget: EUR 263.320, 2014-2017)
- "Computers in Science Education: iCSE" http://icse.us.edu.pl (budget: c.a. EUR 1 milion, funded by EFS, 2011-2014)

He is also co-Author and a task coordinator of PAAD (Platforma Analiz i Archiwizacji Danych) founded by POIG program for 2014-2015 with a total budget c.a. of EUR 4 million. The task "Interactive HPC services for science" main goal is to provide interactive interface to HPC infrastructure (heterogenous cluster of 48 nodes, including 24 GPU and 24 Xeon Phi) using innovative technology of "web notebook" interface.

Publications, products, achievements

1. Sailfish: A flexible multi-GPU implementation of the lattice Boltzmann method Computer Physics Communications Vol.181(9), 2350-2368;2014. Web: http://sailfish.us.edu.pl
2. M.Januszewski and M.Kostur. "Accelerating numerical solution of stochastic differential equations with CUDA", Computer Physics Communications, 181(1):183-188, 2010. Web: https://github.com/marcinofulus/CUDASDE.git
3. iCSE - course materials materials - http://visual.icse.us.edu.pl/iCSE_main/

Previous projects or activities

Internationalization of research and education is one of the priority directions of development of the University of Silesia. The University scientists are actively engaged in research at the international level, actively participates in European Commission initiatives focused both on educational and scientific development, and implements projects within the LLP/Erasmus+ programme, the Research Fund for Coal and Steel, Framework Programmes, as well as the EU Structural Funds.

The institution has been involved in more than 40 FP7 proposals, of which 15 have been funded.

The Faculty of Mathematics, Physics and Chemistry was engaged in the implementation of several FP6 and FP7 projects, i.e.:

1. HadronPhysics (RII3/CT/2003/506078)
2. FlaviaNet (MRTN-CT-2006-035482)
3. LAGUNA (212343)
4. LHCPhenoNet (612536).

There are following projects which are directly connected to infrastructures for virtual research environments:

1. 2011-2014 - iCSE (innovative Computing in Science Education) - grant from European Social Fund, EUR 1 Million, incorporating computational perspective in teaching of mathematics, physics and chemistry using cloud based Sage system and Python language.
2. 2014-30.11.2015 PAAD (Platform for data analysis and archiving) - EUR 3.8Million, funded is mostly HPC center for research with interactive access based on web based notebook UI.
3. 2014-30.11.2015 CNS: Center of Applied Science - 2nd stage, Infrastructure grant includes EUR 500.000 funding for small HPC and cloud infrastructure for education. Technically this will be extension of research HPC center for educational purposes.

Significant infrastructure

The University of Silesia has finished currently executes grants from ESF in total c.a. Million 120 Euro for infrastructure, laboratories and computing centers. New HPC centers are under construction (PAAD and CNS projects) which will provide necessary hardware for development and implementation of virtual research environments.

♠MK [**WRITE HERE**: PIC for University of Silesia]♠
♠**TO DO**: *Reorder partners accordingly to the list on the front page?*♠
[**science-project**]

B.0.4.17 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?*♠

B.0.5 Ethics and Security

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

B.0.5.1 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 authorizing 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♠

B.0.5.2 Security

Please indicate if your proposal will involve:

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

References

[1] http://math.nist.gov/oommf/ (cit. on p. 39).

[2] http://math.nist.gov/oommf/oommf_cites.html (cit. on p. 39).

[3] http://www.ctcms.nist.gov/ rdm/spec3.html (cit. on p. 41).

[4] (Cit. on p. 54).

[5] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, eds. *Handbook of Computational Social Choice*. Cambridge University Press, 2015 (cit. on p. 54).

[6] *CVE-2014-6271*. (cit. on p. 54).

[7] *Docker – An open platform for distributed applications for developers and sysadmin*. hrefhttp://docker.comhttp://docker.com. 2015 (cit. on p. 43).

[8] A. J. Durán, M. Pérez, and J. L. Varona. "The Misfortunes of a Trio of Mathematicians Using Computer Algebra Systems. Can We Trust in Them?" In: *Notices of AMS* 61.10 (2014). , pp. 1249–1252 (cit. on p. 54).

[9] T Fischbacher, M. Franchin, G Bordignon, and H Fangohr. "A systematic approach to multiphysics extensions of finite-element based micromagnetic simulations: nmag". In: *IEEE Transactions on Magnetics* 43.6 (2007), pp. 2896–2898 (cit. on p. 41).

[10] D. Ginev, C. Jucovschi, S. Anca, M. Grigore, C. David, and M. Kohlhase. "An Architecture for Linguistic and Semantic Analysis on the arXMLiv Corpus". In: *Applications of Semantic Technologies (AST) Workshop at Informatik 2009*. 2009. URL: http://www.kwarc.info/projects/lamapun/pubs/AST09_LaMaPUn+appendix.pdf (cit. on p. 78).

[11] M. Iancu, C. Jucovschi, M. Kohlhase, and T. Wiesing. "System Description: MathHub.info". In: *Intelligent Computer Mathematics 2014*. Conferences on Intelligent Computer Mathematics. (Coimbra, Portugal, July 7–11, 2014). Ed. by S. Watt, J. Davenport, A. Sexton, P. Sojka, and J. Urban. Lecture Notes in Computer Science. Springer, 2014, pp. 431–434. URL: http://kwarc.info/kohlhase/submit/cicm14-mathhub.pdf (cit. on p. 78).

[12] C. Jucovschi and M. Kohlhase. "sTeXIDE: An Integrated Development Environment for sTeX Collections". In: *Intelligent Computer Mathematics*. Ed. by S. Autexier et al. LNAI 6167. Springer Verlag, 2010, pp. 336–344. ISBN: 3642141277. arXiv: 1005.5489v1 [cs.OH] (cit. on p. 78).

[13] *Knowledge Adaptation and Reasoning for Content*. URL: http://kwarc.info (visited on 05/12/2011) (cit. on p. 78).

[14] M. Kohlhase. OMDoc *– An open markup format for mathematical documents [Version 1.2]*. LNAI 4180. Springer Verlag, Aug. 2006. URL: http://omdoc.org/pubs/omdoc1.2.pdf (cit. on p. 78).

[15] M. Kohlhase, B. A. Matican, and C. C. Prodescu. "MathWebSearch 0.5 – Scaling an Open Formula Search Engine". In: *Intelligent Computer Mathematics*. Conferences on Intelligent Computer Mathematics (CICM). (Bremen, Germany, July 9–14, 2012). Ed. by J. Jeuring et al. LNAI 7362. Berlin and Heidelberg: Springer Verlag, 2012, pp. 342–357. ISBN: 978-3-642-31373-8. URL: http://kwarc.info/kohlhase/papers/aisc12-mws.pdf (cit. on p. 78).

[16] M. Kohlhase and I. Şucan. "A Search Engine for Mathematical Formulae". In: *Proceedings of Artificial Intelligence and Symbolic Computation, AISC'2006*. Ed. by T. Ida, J. Calmet, and D. Wang. LNAI 4120. Springer Verlag, 2006, pp. 241–253. URL: http://kwarc.info/kohlhase/papers/aisc06.pdf (cit. on p. 78).

[17] M. Kohlhase et al. "The Planetary System: Web 3.0 & Active Documents for STEM". In: *Procedia Computer Science* 4 (2011): *Special issue: Proceedings of the International Conference on Computational Science (ICCS)*. Ed. by M. Sato, S. Matsuoka, P. M. Sloot, G. D. van Albada, and J. Dongarra. Finalist at the Executable Paper Grand Challenge, pp. 598–607. DOI: 10.1016/j.procs.2011.04.063. URL: http://kwarc.info/kohlhase/papers/epc11.pdf (cit. on p. 78).

[18] *MathHub.info: Active Mathematics*. URL: http://mathhub.info (visited on 01/28/2014) (cit. on p. 78).

[19] *MathOverflow*. URL: http://mathoverflow.net (visited on 09/08/2011) (cit. on p. 54).

[20] N. Nisan. “Introduction to mechanism design (for computer scientists)”. In: *Algorithmic game theory*. Cambridge Univ. Press, Cambridge, 2007, pp. 209–241 (cit. on p. 54).

[21] *Nmag — Guided tour*. http://nmag.soton.ac.uk/nmag/0.2/manual/html/guided_tour.html (cit. on p. 42).

[22] *Nmag micromagnetic simulation suite*. http://nmag.soton.ac.uk (cit. on p. 41).

[23] *Notebook based interactive GUI for common simulation requirements*. https://www.youtube.com/watch?v=RhMHgQbP__A. 2014 (cit. on p. 42).

[24] F. Rabe. *The MMT Language and System*. URL: https://svn.kwarc.info/repos/MMT (visited on 10/11/2011) (cit. on p. 78).

[25] F. Rabe and M. Kohlhase. “A Scalable Module System”. In: *Information & Computation* 0.230 (2013), pp. 1–54. URL: http://kwarc.info/frabe/Research/mmt.pdf (cit. on p. 78).

[26] A. Slivkins and J. W. Vaughan. “Online Decision Making in Crowdsourcing Markets: Theoretical Challenges”. In: *SIGecom Exch.* 12.2 (Nov. 2014), pp. 4–23. ISSN: 1551-9031. DOI: 10.1145/2692359.2692364. URL: http://doi.acm.org/10.1145/2692359.2692364 (cit. on p. 54).

[27] *Trac Open Source Project.* (cit. on p. 54).

[28] *VirtualBox*. hrefhttp://virtualbox.orghttp://virtualbox.org. 2015 (cit. on p. 43).