H2020 Call 3 Topic 9-2015

e-Infrastructures for Virtual Research Environments (VRE)

EINFRA-9

Open Digital Research Environment Toolkit for the Advancement of Mathematics

Acronym: OpenDreamKit

H2020 Call 3 Topic 9-2015
Date of Preparation: January 10, 2015

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

Coordinator: Nicolas M. Thiéry

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

tel/fax: +33 6 77 90 32 79

#	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	CNRS Aquitaine	CNRS	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

.

OpenDreamKit will deliver a flexible 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.

By concentrating the efforts on improving and unifying existing general purpose building blocks, OpenDreamKit will simultaneously maximize sustainability and broad impact, with beneficiaries extending to scientific computing, physics, chemistry, biology, engineering, medicine, earth sciences, and geography, and including researchers as well as teachers, and practitioners in the industry.

Contents

1 Excellence

Improvements of the economy, ecology, health care and society overall are driven though innovation. The key enabling tool for innovation advances is mathematics, examples including the global positioning system (GPS) needing relativistic mathematics, and mobile phone connectivity and communication depending on combinatorial optimization and cryptographic algorithms derived from number theory. Engineering, Science and business innovation that enrich society and men kind are made possible due to these mathematical foundations.

Modern mathematical research is increasingly accelerated by and enabled through computational software, such as Sage — an open source mathematics software system — and Jupyter, an open source browser-based notebook with support for code, text, mathematical expressions, inline plots and other media. These tools have the potential to revolutionise the way computational research is conducted.

In this project, we will provide mathematicians and scientists with a generic unified toolkit, the Open Digital Research Environment Toolkit for the Advancement of Mathematics (OpenDreamKit), that allows (i) building of specific Virtual Research Environments (VREs) and (ii) more effective communication of research.

We will achieve this by investing into creation of a *toolkit of software components* from which *tailored VREs can be assembled flexibly* to cater for a variety of needs in maths, science and engineering. We are at a critical point providing an opportunity to do so: emerging collaboration tools for code sharing, such as github, allow to bring together very large sets of open source code developers working on the same codebase. Simultaneously, specialised computational open source tools are emerging. Throughout this project we will use and extend open source code, and OpenDreamKit will benefit from future open source contributions during and beyond the life time of the project. By unifying tools with overlapping functionality, such as Jupyter and Sage, we focus the effort of the computational community onto OpenDreamKit, producing additional economies of scale.

In more detail, VREs based on OpenDreamKit can combine symbolic mathematics, automatic code generation, numerical computation, data bases, post-processing and visualisation in a single document. The document consists of a number of executable cells which contain code, that can be interactively executed, and the output of the code. Cells can also contain arbitrary text and equations as necessary to describe/document the results. These executable documents can be shared with others and fully define and document a computational study – providing step changes in effective research, research communication and reproducibility in computational mathematics and science.

The VREs build on OpenDreamKit will provide end-to-end toolchains that link fundamental mathematics to domain specific specialised computation, this bridging the gap between fundamental research and technology, and paving the way towards faster commercialisation of basic research.

As part of this project, we will also study the social challenges associated with large-scale open source code development, publications based on executable documents, and develop demonstrator VREs based on OpenDreamKit.

The OpenDreamKit team is a Europe-wide collaboration that assimilates a leading body mathematicians and transdisciplinary computational researchers with a track record of delivering innovative open source software solution. All partners are simultaneously code developers and end-users of the toolkit.

By focusing on a toolkit rather than a monolithic VRE, and by concentrating the efforts on improving and unifying existing general purpose building blocks, and in the forefront `Jupyter`, OpenDreamKit will simultaneously maximize sustainability and broad impact. Indeed, even if the primary target users are *researchers in mathematics*, the set of beneficiaries extends to scientific computing, physics, chemistry, biology, engineering, medicine, earth sciences and geography, and include researchers as well as teachers and practitioners in the industry. OpenDreamKit will further foster development models that are mutually beneficial to academia and highly innovative SME's.

1.1 Objectives

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.

Our research will cover a wide variety of aspects, ranging from software development models, user interfaces , deployment frameworks and novel collaborative tools, component architecture, design, and standardization of software 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.

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.

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.

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.

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.

Objective 8: Demonstrate the effectiveness of Virtual Research Environments built on top of OpenDreamKit components for a number of real-world use cases that traverse domains. This addresses part of Aim 2 and through documenting best practice in reproducible demonstrator documents Aim 3.

Objective 9: Effective Dissemination

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. 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.

Our research will cover a wide variety of aspects, ranging from software development models, user interfaces 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 2: 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 community's 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:

Objective 7: Next generation mathematical databases Mathematics has a rich notion of data: it can either be numeric/symbolic data; knowledge about mathematical objects given as statements (definitions, theorems or proofs) or software that computes with these mathematical objects.

All this data is really a commons, and should be maintained as such. Much of this grant proposal and prior work of many of the experts involved is concerned with open source mathematical software, through permissive licensing of their work.

The objective described here is to build infrastructure enabling mathematicians to collaboratively build this commons, while fostering a virtuous cycle of interoperability between these different types of data: one mathematician might implement an algorithm, to be run later on numerical data collected by another scientist.

Objective 8: Collaborative Research Environments that Transcend Domains

We will create a number of demonstrators to highlight the power of OpenDreamKit. The demonstrators will highlight the applicability of OpenDreamKit across mathematics, engineering and science. They will act so as to provide recipes for state-of-the-art computational infrastructure tools and provide avenues for ensuring the repeatability of mathematical analysis.

In particular, we will create a repository of interactive notebooks across a range of application domains (e.g. engineering mechanics, biology and physics). The notebooks will demonstrate a variety of numerical and symbolic techniques in self-contained executable documents. We expect these exemplars to feed in to education at high schools and universities (both undergraduate and postgraduate level). They will also provide a resource for outreach and self-study.

Our demonstrator notebooks will also act as demonstrators of the features developed in OpenDreamKit. They will be incorporated as a developed in this project, can be re-executed as a regression test, and form part of the documentation of OpenDreamKit.

Objective 9:

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.

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.

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 applications by developing a service-driven tool, based on software, knowledge and data integration. Tailored to the researchers' specific needs and workflows, the VRE will support the entire life-cycle of computational work in mathematical research. It will improve the productivity within the community by investigating better collaboration processes, and 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-infrastructure : software development models, collaborative tools, data, component architecture, deployment frameworks, standardization, social aspects, but also fostering collaboration inside the community, community enlargement and links with other scientific communities.
VRE should foster cross-disciplinary data interoperability.	OpenDreamKit will foster a sustainable ecosystem of interoperable source components developed by overlapping communities, and data interoperability between different 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 sharing of rich mathematical data, taking account of provenance and citability. It will allow data sharing in a semantically sound way, and make software sustainable, reusable and easily 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 be seamlessly combined together to build versatile high performance VREs, each tailored to a specific need in pure mathematics and application. We will develop and demonstrate (WP3) a set of APIs 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 etc.). Through well defined APIs, we will enable discovery of subsystems, functionality, documentation 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, including those involving industrial stakeholders. At the end of the project, the effectiveness of the VRE will be demonstrated for a number of real use cases from different domains.
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 and interactive scientific computing environments : `GAP`, `Sage`, `LinBox`, `PARI/GP`, `Singular` and `IPython`, connected to databases, that will allow a huge gain in efficiency and productivity, enabling a large-scale collaboration on software, knowledge, and data.
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 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.

1.3 Concept and Approach

1.3.1 Material previously in the introduction

♠*In short: math is at the heart of innovation*♠

We live in an innovation-driven society. The key enabling tool for our advances is mathematics. For just a few examples, the global positioning system (GPS) needs relativistic mathematics, Computer Assisted Tomography (CAT scanning) is based on solving mathematical inverse problems mobile phone connectivity depends on combinatorial optimization algorithms and Delaunay triangulations for frequency allocation, while the modern communications infrastructure relies on cryptographic algorithms derived from number theory. At the core of each of these innovations there is the underpinning mathematical that is implemented through algorithms. Such innovations have been made possible by investments into pure and applied research in mathematics. Engineering and business innovation is then founded upon these fundamentals to enrich society with the benefits of the mathematical insights..

This proposal is about providing mathematicians with the tools to communicate their research effectively. It will ensure that ideas are distributed and discussed as widely as possible to enable rapid assimilation of these ideas into the pipeline of innovation. Our aim is to develop an ecosystem that exploits the modern computing infrastructure to streamline development and deployment of mathematical advances. We will ensure that the societal benefit of these advances is realized in the shortest possible timescale.

♠Innovation is reciprocally having deep effects on math research, through collaborative and computational tools♠
In the last three decades mathematics research has gone from being a solitary pen-and-paper activity of talented individuals corresponding via lectures, letters, and journal articles to a collaborative, geographically distributed team activity that is supported by e-infrastructures. Simultaneously computational methods have become far more prominent in mathematical research, driven by the availability of computers, high-quality software, and by interest in mathematical problems and proofs that cannot be addressed without massive computer calculations. ♠Possible examples, CoFSG and sequelae, Kepler conjecture, hardware verification.♠

♠*The math community embraces and fosters innovation in technology*♠

The mathematics research community has always been at the core of new technology, from Newton's innovations in reflecting telescopes, to Turing and von Neumann's roles as the founding fathers of Computer Science. The generality of mathematical ideas has been applied to generate important technological advances. In 1945 Alan Turing wrote of his design for the NPL ACE computer, that

> There will be positively no internal alterations to be made even if we wish suddenly to switch from calculating the energy levels of the neon atom to the enumeration of groups of order 720.

. Foreshadowing the dominance of mathematical abstraction (and its implementations in computation) in modern science and technology.

Even in more practical areas, such as in web standards, mathematicians have lead the way. MathML was the first XML recommendation, while `planetmath.org` adopted Web 2.0 standards even before Wikipedia. The theory of high performance computing (HPC) is underpinned by mathematical models of concurrency, as well as a major driver of innovation in domain being Computer Algebra System research. Many now standard programming paradigms appeared originaly there .

♠*A long track record of open source math soft development*♠

Collaborative software innovation by mathematical researchers was key to the development (often in Europe) of many highly successful, open-source, community-developed specialized systems, starting with `PARI/GP` in Number Theory in 1979, and including `GAP` in Group Theory or `Singular` in Algebraic Geometry. This was at a time when most other scientific computing research relied on bespoke Fortran programs used for one calculation and then discarded.

♠*The emergence of massive collaborative development models and tools is revolutionizing the landscape; innovation is now led by communities, not corporate software*♠

In that period inter-project communication was limited by the lack of interconnection between computers, leading to systems remained limited to specific research topics and non-interoperable. It was left to the corporate world to gather sufficient manpower to develop general purpose systems, that could support a broad range of engineering, scientific and statistical mathematics, through a coherent user interface. These companies (e.g. Wolfram, Mathworks, MapleSoft) were mainly US-based and created a profitable industry.

The modern environment, however, is quite different. A more connected digital world has led to the emergence of `Sage`. `Sage` is a truly general purpose computational mathematical system. It is committed to, and draws huge benefits from, the power of open source software a virtual software development environment. It showcases the modern reality that open source software is not just a viable alternative for commercially produced alternatives, but it actually allows for more rapid innovation through providing an open platform through which the community can deploy and share advances more rapidly. `Sage` showcases this modern user-driven community approach to development by delivering high quality software to researchers, teachers, and practitioners in mathematics. It is founded on a widespread international community of contributors and developers and builds successfully on a large stack of existing open source software, ranging from the specialized computational systems mentioned above, to `Python`, a general purpose programming language that is used by millions of programmers worldwide. This flexible, open source architecture then allows it to rapidly assimilate new components such as `Jupyter` (formally IPython notebook) as they are developed.

In the 1980s and 1990s the economies of scale favoured the commercial development model for mathematical software: corporate entities could co-locate a large body of expertise and orient it towards one goal. This was difficult for the much larger, but naturally more dissipated, communities of mathematical researchers. However, modern interconnection of researchers (through the infrastructure of the internet and collaborative development environments such as github) means that the balance of these economies of scale has been reversed. Commercial packages can no longer develop fast enough to assimilate the innovation of the wider mathematical community, where there is greater expertise and manpower.

This proposal is about supporting the next generation of innovation in the mathematical computing ecosystem. It is a Europe-wide collaboration that assimilates a leading body mathematicians and computational researchers with a track record of delivering innovative open source software solutions.

> ♠In short: maths needs VRE's♠
> *Our priority is the delivery of complete Virtual Research Environments (VRE). A VRE supports the entire life-cycle of computational work in mathematical research, from initial exploration to publication, teaching, and outreach. We envisage VREs as the main medium for development and deployment of mathematical research.*

Virtual Research Environments are flexible, powerful, unified environments for communication, distribution and implementation of mathematical research. Initial work shows the potential for this idea, for example, the Virtual Research and Teaching Environment `SageMathCloud` hosting more than 10k users and 100k projects after just one year). There is widespread community interest in well-executed *integrated solutions* which can enable large-scale collaboration on Mathematical *software*, *knowledge*, and *data*. This interest is also evidenced by the considerable activity (since the inception of the internet) in a range of online mathematical databases such as the Online Encyclopedia of Integers Sequences, the Atlas of Group Representations, and `LMFDB`. Other systems such as polymath and MathOverflow show the interest among mathematicians in exploring new forms of collaboration, in particular when the tools are well-designed and the balance of effort and reward is correct.

♠*Challenges*♠

Engineering the social aspects of such systems to maximize their success is an imprecise science. A great deal is learnt from the deployment of any given system and the reaction of the wider community. Historically setting up these infrastructures has required massive ad-hoc efforts: `SageMathCloud` required 70k lines of bespoke code to integrate its components and each of the databases and collaboration sites is essentially a bespoke program. Unfortunately much of the effort is necessarily *not* on the innovation in the environment, but on the underlying infrastructure. To ensure focus can be placed on the environment itself we require a more flexible and reusable system. Its characteristics must include portability, compatibility, performance, usability, and reproducibility. It should bootstrap our understanding of the social dynamics of user and developer communities.

♠*Specific challenges in maths*♠

A specific challenge in mathematics comes from the vast yet tightly connected array of concepts involved. The natural ontologies of mathematics are richer, more complex, and more interconnected than for, say, fluid dynamics. Mathematical collaborations and mathematical research projects involving collaboration also vary widely, as do the skills and needs of individual mathematicians, calling for a highly modular and customizable VRE infrastructure.

♠Our approach♠
*OpenDreamKit proposes to deliver a flexible **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.*

In keeping with the `Sage` strategy, a major focus is on reusing and improving existing components, and reaching toward larger communities whenever possible. A key technology here is `Jupyter` (formerly `IPython`), a set of open-source software projects for interactive and exploratory computing. These software projects help make scientific computing and data science reproducible and multilanguage (`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, text, interactive dashboards and other rich media into a single executable document. `Jupyter` is already used very widely in research and development, both in academia and industry and the user base grows rapidly.

♠Short description of the consortium♠
To achieve this aim, OpenDreamKit brings together lead developers and experts from existing mathematical computational components (LinBox, GAP, Sage, Singular), existing Virtual Research Environments (SageMathCloud, Simulagora), online mathematical databases (LMFDB), mathematics knowledge portals (MathHub) and general purpose interactive computing components (Jupyter).

♠Benefits of the approach: sustainabiblity and massive impact♠
By focusing on a toolkit rather than a monolithic VRE, and by concentrating the efforts on improving and unifying existing general purpose building blocks, and in the forefront Jupyter, OpenDreamKit will simultaneously maximize sustainability and broad impact. Indeed, even if the primary target users are researchers in mathematics*, the set of beneficiaries extends to scientific computing, physics, chemistry, biology, engineering, medicine, earth sciences and geography, and include researchers as well as teachers and practitioners in the industry. OpenDreamKit will further foster development models that are mutually beneficial to academia and highly innovative SME's.*

Overall concept The ambition of this project, set out below, is to develop tools and techniques that will allow individual researchers or groups to compose a VRE which will provide modern, flexible and reliable support for the whole lifecycle of a mathematical research project, including collaborative exploration, mathematical computations of all scales from tiny to huge, proof, publication and archival. To do this they will need a toolkit from which to assemble Digital Research Environments for the Advancement of Mathematics – the DREAMKit.

The kit will take the form of a collection of compatible components, ready to be connected using extensible documented interfaces both to other bespoke components and to standard infrastructural tools and services. Most of the capabilities of these components will come from existing software – computational tools such as `Sage`, `GAP`, `Singular` and `PARI`; databases such as LMFDB; user interface tools such as the iPython notebooks (now part of the `Jupyter` project); existing compute servers, clusters and clouds; typesetting tools such as LaTeX and so on. Some of these need to be updated to run to best effect in modern environments, most will need adaptation to support the OpenDreamKit interfaces. Some new components may be needed. Groundwork must also be laid for effective consistent maintenance and future development of these components.

The kit will be designed to create VREs that support the ways in which mathematicians actually work together through the lifecycle of a project, informed by recent research into the sociology of mathematical collaboration.

The activities of the project are planned and structured to develop and promote the OpenDreamKit, including new research into architectures, database techniques, parallel algorithms and the sociology of collaborative free software development, as well as engineering work on existing software and networking and community-building activities.

The project inherently spans the disciplines of mathematics and computer science, as well as bringing in results and techniques from social sciences. Exemplar applications may also arise from areas to which symbolic and algebraic computing is applied, such as physics, chemistry, systems biology and engineering.

Approach and methodology The project is divided into seven work packages. Work Package **??** covers project management and coordination as usual. Work Package **??** is our main Networking activity including community-building workshops, demonstrator applications and direct dissemination of project results. This covers the following topics from section E of the Work Programme:

- dissemination and/or exploitation of project results and knowledge, contribution to socioeconomic impacts, promotion of innovation ;
- reinforcing partnership with industry: outreach and dissemination activities, transfer of knowledge, activities to foster the use of e-infrastructures by industrial researchers, involvement of industrial associations in consortia or in advisory bodies;
- strengthening of virtual research communities;
- spreading of good practices, consultancy and training courses to new users;
- exchange of personnel and training of staff;

The remaining work packages are Joint Research Activities, dividing up the research needed to design and implement the OpenDreamKit and investigate the best models for its future development.

This covers the following topics from section E of the Work Programme:

- higher performance methodologies and protocols, higher performance instrumentation, including the testing of components, subsystems, materials, techniques and dedicated software;
- integration of installations and infrastructures into virtual facilities;
- innovative solutions for data collection, management, curation and annotation;

Additional topics addressed include effective software development and maintenance methodologies for systems of free software systems and the design of VREs to best support real mathematical practice.

Since the infrastructure that we are developing is free software, there is no need for formal Service activities. All partners have access to all the software anyway, and development and demonstration can take place on computers already available to the partners.

Networking activities

Joint Research Activities

Use of Existing Basic Services

Gender analysis All partners will follow inclusive practices in recruiting staff for this project, in inviting the community to our workshops and outreach events and in choosing users to evaluate our demonstrator applications. We will consult with about any known gender differences in collaborative working and ensure that our collaborative tools properly support open, equitable and inclusive patterns of cooperation. .

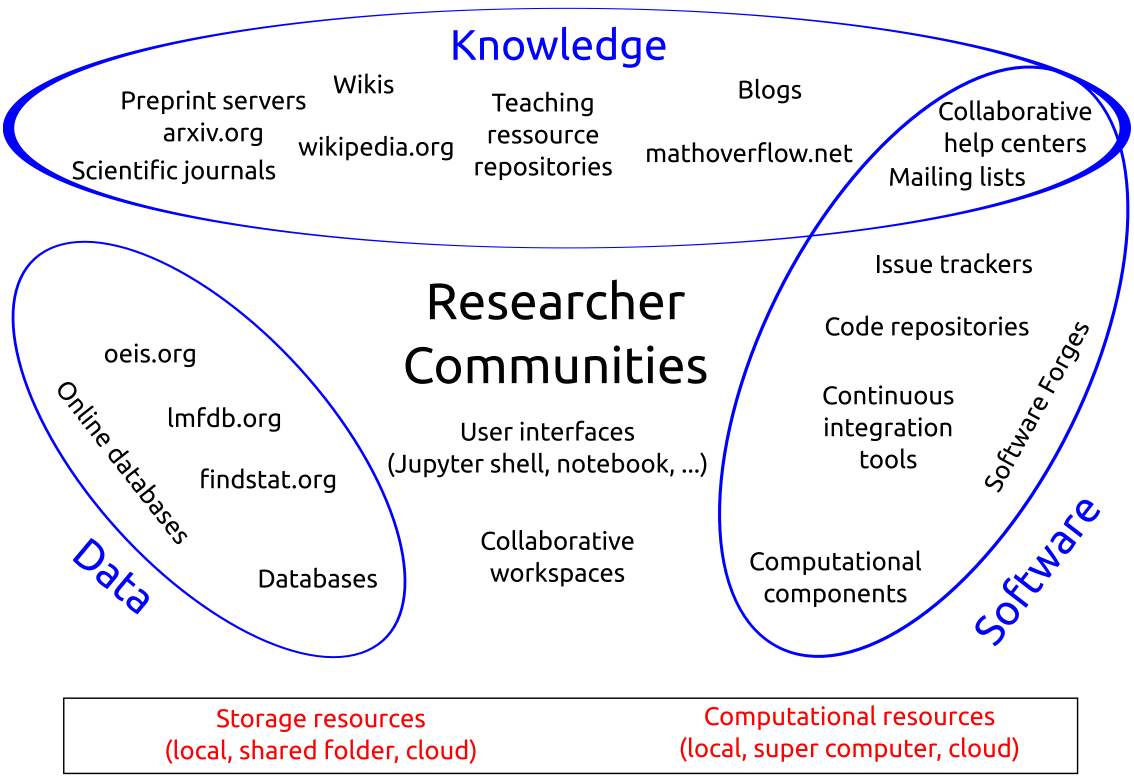

Figure 1: Virtual Research Environments for research in pure mathematics and applications.

Today's research is transformed by the Internet and the availability of vast amounts of research data on the Internet in virtual research environments. Arguably, 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:

- *$\mathcal{D}$: proper (numeric/symbolic) data*
- *$\mathcal{K}$: the knowledge about the mathematical objects given as statements (definitions, theorems or proofs; either formal or rigorously informal)*
- *$\mathcal{S}$: software that computes (with) the mathematical objects*

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

- *$\mathcal{D}$ serves as examples for $\mathcal{K}$ or as counterexamples for conjectures in $\mathcal{K}$;*
- *$\mathcal{S}$ computes $\mathcal{D}$ and establishes properties of $\mathcal{D}$ (given as $\mathcal{K}$);*
- *$\mathcal{D}$ tests $\mathcal{S}$, $\mathcal{S}$ is verified with respect to $\mathcal{K}$;*
- *theorems and proofs in $\mathcal{K}$ induce and justify algorithms for $\mathcal{S}$;*
- *$\mathcal{D}$ induces conjectures and guides proofs in $\mathcal{K}$.*

Figure 1 instantiates this situation with respect to the $\mathcal{DKS}$-resources that are already in use in Mathematics. We name just a few paradigmatic systems that are relevant in the scope of the OpenDreamKit project:

1. **Data Repositories/Communities**: Many communities have been collecting and sharing data about the objects they study: e.g.

   a. The *Open Encyclopedia of Integer Sequences* [`http://oeis.org`] has collected sequences of integers for half a century, it now contains publications about, relations between, programs for, and data on ca. 250.000 sequences and is steadily growing

   b. The *database of L-Functions, Modular Forms, and related objects* (LMFDB) is an extensive database of mathematical objects arising in Number Theory. The associated website aims to become a modern handbook including tables, formulas, links, and references, to these objects, including specific L-functions and their sources.

c.

2. **Knowledge Sources and Repositories** There are many ways to represent mathematical knowledge and involve computers. Systems and resources range from relatively traditional pre-publication systems like

   a. the *Cornell EPrint archive* [http://arxiv.org] has over 1 million LaTeX-based pre-prints of which ca 10-15% are on mathematics and bordering areas.
   b. via community-driven Q/A sites like [http://mathoverflow.net] with almost 40 thousand questions answered
   c. to mathematical encyclopedias like [http//planetmath.org], which as a Web2.0 site predates Wikipedia,
   d. the LMFDB website [http://www.lmfdb.org] which includes novel ways to present this data, following a principle called *transclusion*, and in the extreme to
   e. formalizations of mathematical knowledge, e.g. in theorem prover libraries like Mizar [http://mizar.org], which has formalized 50 thousand relatively elementary theorems in 40 years or the formalizations of the Feit-Thomson Theorem or the Kepler Conjecture.

3. **Mathematical Software Development and Systems**

   a.

OpenDreamKit aims to create a framework to make the systems interoperable and synergistic and to give working mathematicians full access to the potential spanned by already-existing systems. Essentially every node in Figure 1 represents a user community, so OpenDreamKit is at its heart a project that also combines researchers and communities.

1.3.2 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 (), 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.

1.3.3 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.

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.

Mathematicians have a strong tradition of sharing knowledge openly (arxiv, Wikipedia, ...).

1.3.4 Early VRE's

1.3.5 Key concept: bringing communities together toward a VRE kit

1.3.6 Linked research and innovation activities

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`

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 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/

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

Logilab: simulagora, cubicweb, ...

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 http://www2.warwick.ac.uk/fac/sci/maths/people/staff/john_cremona/lmf

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 (grant reference EP/K034383/1) 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. 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. It also implements "Knowls", a very fruitful method of presenting mathematical knowledge.

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.

Edith Elkind's ERC Starter Grant awarded in 2014, titled "Algorithms for Making Complex Decisions on Structured Domains", will develop theoretic tools for analysing and improving situations arising in collaborative environments. It can be viewed as a interdisciplinary project, bringing together methods from computer science, game theory, and economics and political science to quantify complex behaviour of social interactions.

OpenDreamKit appears to be a natural testing ground and a potential virtual laboratory for developing and testing ideas and tools developed, within the framework of the ERC Grant, on in a "real life" situation, and the collaboration will be mutually beneficial for both projects.

Ursula Martin's EPSRC funded project "MathSoMac: The Social Machine of Mathematics" (EP/K040251/2) brings rigorous methods from social sciences into studying of the crowdsourcing, e.g. large-scale online collaboration, phenomenon in mathematical sciences.

OpenDreamKit and VREs in general are natural objects to investigate for the latter project, and conclusions drawn would lead to better understanding of the ways VREs function. This has important potential benefits for OpenDreamKit, and vice versa.

Findstat?

KWARC group

HPCGAP

CoDiMa is a new EPSRC funded Collaborative Computational Project in the area of *Co*mputational *Di*screte *Ma*thematics (EP/M022641/1). It will begin in 2015 and will be aimed at GAP and Sage community-building activities in the UK, involving a programme of short research visits, workshops and training events. Through CoDiMa, we will have an excellent opportunity to interact with UK user and developer communities of GAP and Sage in order to to collect feedback about their requirements and to inform them about OpenDreamKit outcomes.

1.4 Ambition

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.

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: 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.

1.4.1 Challenges specific to mathematics

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

- 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
- interlinking very high
- several alternate and defining description of same objects

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.

2 Impact

2.1 Expected Impacts

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-infrastracture 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%

<table>
<tr><th colspan="2">Higher efficiency and creativity in research, higher productivity of researchers thanks to reliable and easy access to discovery, access and re-use of data</th></tr>
<tr><td>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.</td><td>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</td></tr>
<tr><th colspan="2">Accelerated innovation in research via an integrated access to digital research resources, tools and services across disciplines and user communities</th></tr>
<tr><td>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.</td><td>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</td></tr>
<tr><th colspan="2">Researchers able to process structured and qualitative data in virtual and/or ubiquitous workspaces</th></tr>
<tr><td>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.</td><td>Key performance indicators :
Note E.S. – Nicolas, ici, j’ai du mal, car pas sure si on répond vraiment à ce critère.</td></tr>
<tr><th colspan="2">Increased take-up of collaborative research and data sharing by new disciplines, research communities and institutions</th></tr>
<tr><td>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.</td><td>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</td></tr>
</table>

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)

2.1.3 Other impacts (environmental and socially important impacts)

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.

2.2 Measures to Maximise Impact

2.2.1 Dissemination and Exploitation of Results

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 international conferences like FPSAC, ISSAC, ... (4), international congress of mathematical software (1) 5. Workshops "Sage and women" (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.

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).

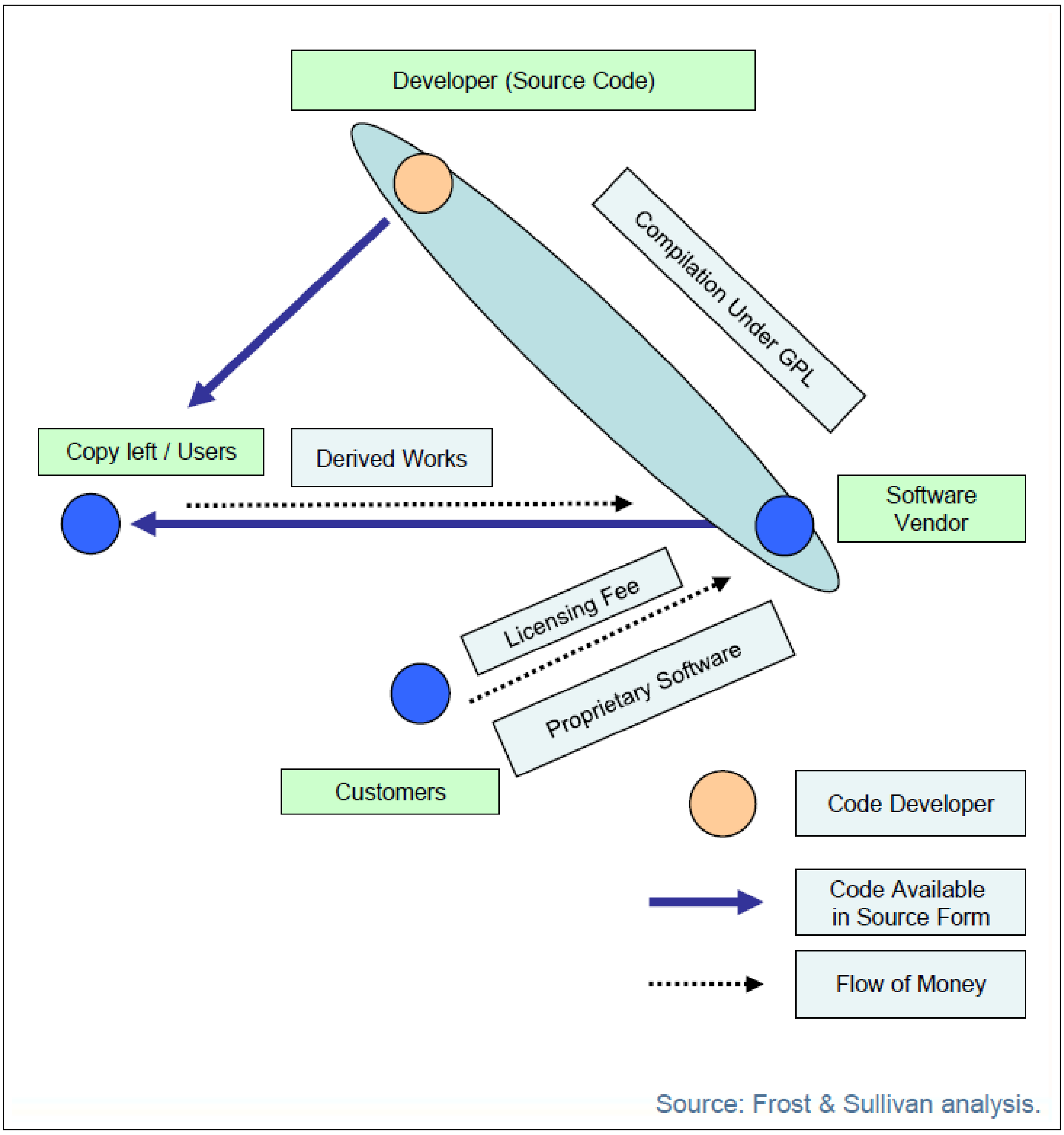

Figure 2: The GPL Model

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.

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.

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.

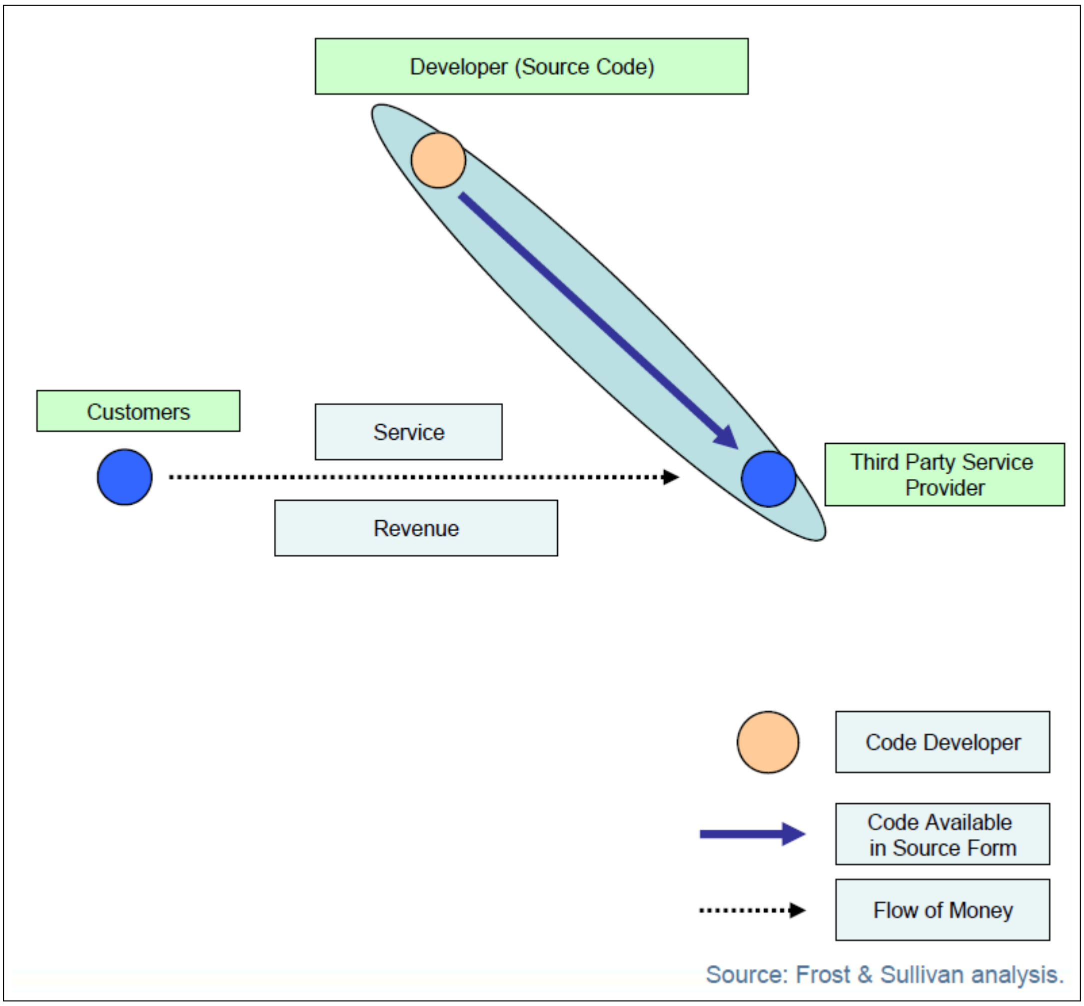

Figure 3: The Third-Party Services Model

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 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.

2.2.2 Communication activities

3 Implementation

3.1 Work Plan — Work packages, deliverables and milestones

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 (WP2). The Gantt chart on Page 27 illustrates the timeline for the various tasks for these work packages, including inter-task dependencies.

How the Work Packages will Achieve the Project Objectives

The project objectives (Section 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

WP	Title	UPS	Logilab	UVSQ	UJF	CNRS	UO	USHEF	USO	USTAN	UW	JacU	UK	US	UZH	Simula	UWS	total
WP1	Management	*28*	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	**58**
WP2	Community Building/Dissemination	*15*		2		4		8	10	18				24		1		**82**
WP3	Component Architecture	64		*8*				4	6	16								**98**
WP4	User Interfaces	26	1	2		1		12	16	18		22	1			*23*		**122**
WP5	High Performance Mathematical Computing	6	12		*52*	1				18			1					**90**
WP6	Data/Knowledge/Software-Bases	4	2					12		10	25	*36*			12			**101**
WP7	Social Aspects						*53*	8	6									**67**
totals		**143**	**17**	**14**	**54**	**8**	**55**	**46**	**40**	**82**	**27**	**60**	**4**	**26**	**14**	**26**	**2**	**618**

Efforts in PM; WP lead efforts light gray italicised

Figure 5: Work Packages

Deliverables

#	Deliverable name	WP	Lead	Type	Level	Due
D1.1	Consortium Agreement	WP1	UPS	R	CO	1
D1.6	Establishing basic project infrastructure (websites, wikis, issue trackers, mailing lists, repositories)	WP1	UPS	DEC	PU	1
D5.1	Facility to compile `Pythran` compliant user kernels and sage code.	WP5	??	DEM	PU	3
D5.2	Parallelise the relation sieving component of the Quadratic Sieve	WP5	??	DEM	PU	3
D5.16	Turn the Python prototypes for tree exploration into production code, integrate to `Sage`.	WP5	??	DEM	PU	3
D3.11	Understand and document `SageMathCloud` backend code.	WP3	??	R	PU	6
D3.1	One-click install `Sage` distribution for Windows with Cygwin 32bits	WP3	??	OTHER	PU	6
D3.7	Virtual images and containers	WP3	??	OTHER	PU	6
D4.1	Active Documents based on sTeX	WP4	JacU	R	PU	6
D3.2	Python Interface to micromagnetic OOMMF package completed	WP3	USO	OTHER	PU	9
D4.2	Full-text search (formulae + Keywords) over LaTeX-based documents (e.g. the arXiv subset)	WP4	JacU	OTHER	PU	9
D5.4	Improve `Pythran` runtime support to automatically take advantage of multi-cores and SIMD instruction units and use it in `Cython`.	WP5	??	DEM	PU	9
D5.5	Implement a parallel version of Block-Wiederman linear algebra over GF2 and the triple large prime variant.	WP5	??	DEM	PU	9
D6.1	Workshop Report	WP6	JacU	R	PU	9
D6.2	Heuristic Parser for the OEIS	WP6	JacU	OTHER	PU	9
D1.3	Internal Progress Reports, including risk management and quality assurance plan	WP1	UPS	R	CO	12
D2.3	Demo: Problems in Physics with Sage v1	WP2	??	DEM	PU	12
D3.5	Support for the SCSCP interface protocol in all relevant components (`Sage`, `GAP`, etc.) distribution	WP3	USTAN	OTHER	PU	12
D4.11	Tools for collaborating on notebooks via version-control	WP4	Simula	OTHER	PU	12
D4.3	Distributed, Collaborative, Versioned Editing of Active Documents in MathHub.info	WP4	JacU	DEM	PU	12
D4.4	Basic `Jupyter` interface for GAP, Pari, `Sage`, Singular	WP4	??	OTHER	PU	12
D4.5	Full featured `Jupyter` interface for GAP, Pari, Singular	WP4	??	OTHER	PU	12
D4.6	`Sage` notebook / `Jupyter` notebook convergence	WP4	??	DEM	PU	12
D5.10	Extend the existing assembly superoptimiser for AVX and upcoming Intel processor extensions	WP5	??	DEM	PU	12
D5.11	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
D5.16	Explore the possibility to interface smoothly `Pythran`, `Cython` and `Cilk++`.	WP5	??	DEM	PU	12

#	Deliverable name	WP	Lead	Type	Level	Due
D5.13	Library design and domain specific language exposing LinBox parallel features to Sage	WP5	UJF	R	PU	12
D5.6	Make Pythran typing better to improve error information.	WP5	??	DEM	PU	12
D5.7	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
D5.8	Parallelise the Singular sparse polynomial multiplication algorithms	WP5	??	DEM	PU	12
D5.9	Parallel versions of the Singular sparse polynomial division and GCD algorithms.	WP5	??	DEM	PU	12
D6.3	Design of Triform (DKS) Theories (Specification/RNC Schema/Examples)	WP6	JacU	R	PU	12
D6.4	Python/Sage Syntax Foundation Module in OMDoc/MMT	WP6	JacU	DEC	PU	12
D6.5	Conversion of existing and new databases to unified interoperable system	WP6	UZH	DEC	PU	12
D6.6	LMFDB deep modelling: Fragment Identification & Initial Model Design	WP6	UZH	R	PU	12
D4.7	Micromagnetic VRE software completed	WP4	USO	DEM	PU	15
D1.6	Innovation Management Plan	WP1	UPS	R	CO	18
D2.10	Short Course: A short course for lecturers on using OpenDreamKit for delivering mathematical education.	WP2	??	DEC	PU	18
D4.10	Facilities for running notebooks as verification tests	WP4	Simula	OTHER	PU	18
D4.8	Full-text search (Formulae + Keywords) over Notebooks	WP4	JacU	OTHER	PU	18
D4.9	In-place computation in active documents (context/computation)	WP4	JacU	DEM	PU	18
D5.14	Report on development of designs for the GAP developments – parallel library, interacts to standard infrastructure and Cython-like extensions	WP5	USTAN	R	PU	18
D6.7	LMFDB Data vs. Knowledge vs. Software Validation	WP6	UZH	R	PU	18
D6.8	Cross-Validation for OEIS DKS-Theories	WP6	JacU	R	PU	18
D4.12	Micromagnetic VRE tutorial and documentation notebooks	WP4	USO	DEC	PU	21
D2.8	Community-curated indexing tool (open source)	WP2	??	P	PU	24
D2.9	Community-curated indexing service for OpenDreamKit	WP2	??	DEM	PU	24
D3.12	*Personal* SageMathCloud: single user version of SageMathCloud distributed with Sage.	WP3	??	OTHER	PU	24
D3.13	Integration between SageMathCloud and Sage's TRAC server	WP3	??	OTHER	PU	24
D3.3	One-click install Sage distribution for Windows with Cygwin 64bits	WP3	UPS	OTHER	PU	24
D3.8	Open package repository for Sage	WP3	??	OTHER	PU	24

#	Deliverable name	WP	Lead	Type	Level	Due
D4.13	Notebook Import into MathHub.info (interactive display)	WP4	JacU	DEM	PU	24
D4.14	Demonstrator online portal available	WP4	USO	DEC	PU	24
D4.15	`Jupyter` extension for 3d CFD visualization	WP4	??	OTHER	PU	24
D4.18	Refactorization of `Sage`'s `Sphinx` documentation system	WP4	??	OTHER	PU	24
D5.15	Ongoing support of Intel, AMD, ARM, Sparc processors and new Operating System versions	WP5	??	DEM	PU	24
D5.16	Refactor the existing combinatorics `Sage` code using the new developed `Pythran` and `Cython` features.	WP5	??	DEM	PU	24
D5.17	Algorithms and implementations. Library maintenance and close integration in mathematical software	WP5	UJF	DEM	PU	24
D6.10	Implementation of Triform Theories in the MMT API.	WP6	JacU	OTHER	PU	24
D6.11	Python/Sage Computational Foundation Module in OMDoc/MMT	WP6	JacU	OTHER	PU	24
D6.9	Shared persistent memoization library for Python/Sage	WP6	UZH	OTHER	PU	24
D7.2	Demonstrator: Mechanism for comment on posted Jupyter notebooks.	WP7	USHEF	DEM	PU	24
D2.4	Demo: Problems in Physics with Sage v2	WP2	??	DEM	PU	30
D4.16	Formula search in CAS programs and Software Modules	WP4	JacU	OTHER	PU	30
D2.5	Micromagnetic VRE code and documents source online	WP2	USO	DEC	PU	32
D1.5	Internal Progress Reports, including risk management and quality assurance plan	WP1	UPS	R	CO	36
D2.11	Demo: Interactive lecture notes and marking systems based on OpenDreamKit.	WP2	??	DEM	PU	36
D2.2	Demo: Nonlinear Processes in Biology interactive book	WP2	??	DEM	PU	36
D3.4	Continuous integration platform for multi-platform build/test.	WP3	??	DEM	PP	36
D3.6	Semantic-aware `Sage` interface to `GAP`.	WP3	UO	OTHER	PU	36
D4.17	Exploratory support for live notebook collaboration	WP4	Simula	OTHER	PU	36
D4.19	Exploratory support for semantic-aware interactive widgets providing views on objects represented and or in databases	WP4	??	DEM	PU	36
D5.18	Implementations of algorithms using distributed memory et heterogenous architectures: clusters and accelerators	WP5	UJF	DEM	PU	36
D6.12	Python/Sage Declarative Semantics in OMDoc/MMT	WP6	JacU	OTHER	PU	36
D6.13	`LMFDB` Algorithm verification wrt. a Triformal theory	WP6	JacU	OTHER	PU	36
D7.2	Demonstrator: Jupyter Notebook Live Poster	WP7	USHEF	DEM	PU	36
D2.2	Demo: Classical Mechanics interactive book	WP2	??	DEM	PU	40
D4.20	Search from Notebooks/Active Documents Interface	WP4	JacU	OTHER	PU	42

#	**Deliverable name**	**WP**	**Lead**	**Type**	**Level**	**Due**
D2.6	Demo: Problems in Physics with Sage v3	WP2	??	DEM	PU	44
D2.7	Micromagnetic VRE workshops delivered	WP2	USO	OTHER	PU	44
D1.6	Innovation Management Plan	WP1	UPS	R	CO	45
D6.14	LMFDB full integration of algorithms, data and presentation (not so much verification)	WP6	JacU	R	PU	46
D5.19	Implementations of the GAP developments, ready for release	WP5	USTAN	OTHER	PU	47
D3.10	HPC enabled Sage distribution	WP3	??	OTHER	PU	48
D3.9	Packaging for major Linux distributions	WP3	??	OTHER	PU	48
D5.20	final report and evaluation of the GAP developments	WP5	USTAN	R	PU	48
D7.3	Micromagnetic VRE environment evaluation report	WP7	USO	R	PU	48

Milestones

The work in the OpenDreamKit project is structured by seven milestones, which coincide with the project meetings. 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.

<table>
<tr><td>#</td><td>Name</td><td>WPs[1]/Deliverables involved</td><td>Mo</td><td>Means of Verif.</td></tr>
<tr><td>M1</td><td>Kickup meeting</td><td></td><td>1</td><td>meeting report</td></tr>
<tr><td colspan="5">Set up the organizational infrastructure, in particular: Web Presence, project TRAC,. . .</td></tr>
<tr><td>M2</td><td>Consensus</td><td></td><td>24</td><td>Inspection</td></tr>
<tr><td colspan="5">Reach Consensus on the way the project goes</td></tr>
<tr><td>M3</td><td>Exploitation</td><td></td><td>36</td><td>Inspection</td></tr>
<tr><td colspan="5">The exploitation plan should be clear so that we can start on this in the last year.</td></tr>
<tr><td>M4</td><td>Final Results</td><td></td><td>48</td><td>Inspection</td></tr>
<tr><td colspan="5">all is done</td></tr>
</table>

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

3.2 Work Package Descriptions

Work Package 1: Project Management																	**Start:** 0
Site	*UPS*	Logilab	UVSQ	UJF	CNRS	UO	USHEF	USO	USTAN	UW	JacU	UK	US	UZH	Simula	UWS	**all**
Effort	*28*	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	**58**

Objectives

Establish and maintain an effective contract, project, and operational management approach, ensuring (i) an effective and timely implementation of the project, (ii) quality control of the results, (iii) risk and innovation management of the project as a whole, as well as (iv) timely and necessary interaction with the EC and other interested parties.

Description

The project will be managed by UPS, which has profound experience in administrating and leading EU funded and national projects. The coordinator together with the WP leaders, will be responsible for monitoring WP status, coordination of work plan updates and annual internal progress reports. The project management structure and roles of partners in the consortium are presented in the next section.

T1.1 Project and financial management Sites: UPS (lead)

The task includes the following activities

- Preparation and Distribution of the Consortium Agreement;
- Setup project website, intranet and communication procedures for effective communication;
- Organization of project review and progress meetings;
- Establishment and maintenance of external contacts (with the EC, other relevant national / EU projects, other academic and industrial stakeholders) to organize transfer of knowledge, present and promote project results;
- Progress and Financial Reporting to the EC;
- Data and IPR Management will be managed in accordance with agreed rules stated in the Consortium Agreement and in accordance with the Data management plan (Task X.Y).

T1.2 Quality assurance and risk management Sites: UPS (lead)

A quality assurance plan will be established to ensure coherent and sufficient quality of the work and results. The plan will be developed by UPS, involving all partners, and will be followed up regularly. In addition, the project coordinator with support from the coordination team and quality review board will establish and review annually a risk management plan and self-assessment to ensure that technical barriers / potential risks are identified and corrective measures are put into place on time (D1.3).

T1.3 Innovation management Sites: UPS (lead)

One of the most important criteria for success for the OpenDreamKit project is to bring the project results into use. Therefore, exploitation routes will be sought whenever possible. In order to create a common understanding within the Consortium about how we can best usher an idea all the way from conception to its realization and exploitation, the Coordinator will be responsible for the preparation and realization of an Innovation Plan to assure that research activities meet the required milestones and the outputs of innovation are fully aligned with the project objectives. All research activities will go through three initial steps where the exploitation opportunity is identified along with the main stakeholders for the exploitation opportunity and an IP owner.

Deliverables:

D1.1 (Due: 1, Type: R, Dissem.: CO, Lead: UPS) *Consortium Agreement*

D1.2 (Due: 1, Type: DEC, Dissem.: PU, Lead: UPS) *Establishing basic project infrastructure (websites, wikis, issue trackers, mailing lists, repositories)*

D1.3 (Due: 12, Type: R, Dissem.: CO, Lead: UPS) *Internal Progress Reports, including risk management and quality assurance plan*

D1.4 (Due: 18, Type: R, Dissem.: CO, Lead: UPS) *Innovation Management Plan*

D1.5 (Due: 36, Type: R, Dissem.: CO, Lead: UPS) *Internal Progress Reports, including risk management and quality assurance plan*

D1.6 (Due: 45, Type: R, Dissem.: CO, Lead: UPS) *Innovation Management Plan*

Work Package 2: Community Building, Training, Dissemination, Exploitation, and Outreach								**Start:** 18
Site	*UPS*	UVSQ	CNRS	USHEF	USO	USTAN	US	Simula	**all**
Effort	*15*	2	4	8	10	18	24	1	**82**

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. This includes:

- reviewing emerging technologies;
- ensuring awareness of the results in the user community;
- engaging cross communities discussions to foster scientific collaborations and conjoint developments;
- spreading the expertise through workshops and trainings;
- defining individual exploitation plans; and,
- managing existing and new intellectual property.

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. This is also an occasion to organize cross communities workshops like Sage-Jupyter days.

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.

T2.1 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.

T2.2 Dissemination and Communication activities

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), 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.

T2.3 Community building: development workshops Sites: UPS (lead)

We will organize development workshops all throughout the project. The aim of these workshops is to bring together developers from the different communities to implement some key aspects of OpenDreamKit: user interface, documentation, cross compatibility, etc. These meetings will gather not only participants of OpenDreamKit but also members of the different communities involved. Bringing talented people together is the best way to make actual progress on the different aspects of OpenDreamKit. It is also a way to work within the communities we're reaching and include them in the discussions and development. It fosters collaborations between scientists and developers from different backgrounds to build tools that are needed by all.

Each workshop will be aimed at a specific software (`Sage`, `GAP`, `SageMathCloud`, `IPython`, `Singular`, etc.) or be joint meeting between different communities to improve interoperability and joint developments. We are planning to have 4 or 5 such events per year and will produce a activity report every year. More precisely, here is a tentative list of the different events planned so far:

- One Sage worshop per year in Cernay France (near Orsay) where similar gatherings took place before. One of them will be a Sage-Sphinx day, dedicated to documentation.
- One Atelier PARI in Bordeaux per year. The team in Bordeaux has a great experience in organizing this kind of PARI events.
- Two Singular workshops and two GAP workshops in Kaiserslautern over the four years.
- Two workshops dedicated to high performance mathematical computing in relations with WP5. One of them should be in Grenoble and the second one in Bordeaux to foster the work with PARI towards WP5.**T5.1**.
- A joint meeting on the topics of SageMathCloud and Jupyter in Simula in relations with WP4.
- A joint event between GAP, Sage, and Singular in ICMS, Edinburgh.
- A joint Jupyter and Sage event in Orsay.
- A joint LMFDB and Sage event in Warwick to work towards WP6.

T2.4 Introduce OpenDreamKit to researchers and teachers

In this task, we will develop and deliver materials that will introduce OpenDreamKit to potential users—both researchers and teachers.

Develop a 'taster' seminar (1-2 hours) and follow-up short course (1-2 days) on OpenDreamKit for researchers and lecturers in all disciplines D2.10. At Sheffield, this will be added to the set of courses that are offered as part of IT Services' research support department. As such, it could potentially reach all disciplines. It will also be made publicly available for widespread dissemination and collaborative modification.

Elements of this work will also be integrated with the GP Summer Schools and Roadshows. The Summer School is now in its fourth edition (over 140 students educated). The Roadshows have taken place in Uganda, Colombia and are scheduled for Italy, Australia and Kenya.

These seminars and short-courses will be used to identify potential collaborators who are interested in utilising OpenDreamKit immediately. We will act as consultants to these collaborators in two ways:

We will work with lecturers at Sheffield to introduce OpenDreamKit to various disciplines via the production of interactive lecture notes D2.11. The focus for the student here will not necessarily be on programming but rather on interaction with the subject matter via use of OpenDreamKit. Interactive lecture notes are an area where commercial vendors such as MapleSoft and Wolfram Research are spending a lot of time and money developing material. We will provide technical and programming expertise to lecturers—helping them to develop the interactive part of notes while they provide the subject material.

We will work as consultants with researchers at Sheffield to introduce OpenDreamKit to their workflow. Any projects that successfully do this will be promoted as case studies for OpenDreamKit.

T2.5 Open source dissemination of micromagnetic VRE Sites: USO (lead)

Tasks WP4.**T4.8** and WP4.**T4.9** provide a micromagnetic VRE demonstrator built on top of OpenDreamKit. In this task, we set up of the infrastructure (D2.5) to encourage and invite code contributions from the micromagnetic community to both code and created notebooks, while automating quality control and maintaining trust effectively.

The source code of the micromagnetic VRE will be made available as open source on public repository hosting sites (such as GitHub/Bitbucket), and announced to the community via appropriate mailing lists and other means. We will set up a publicly accessible Jenkins/Travis continuous integration (CI) system to (i) run regression tests (from WP3.**T3.8** and WP4.**T4.8**) routinely when the micromagnetic VRE code or underlying OOMMF core code changes, (ii) re-execute notebooks (from WP4.**T4.9**) and use them as regression tests (using the outcome of task WP4.**T4.3**), and (iii) re-build downloadable installation files and virtual machine images.

T2.6 Micromagnetic VRE dissemination workshops Sites: USO (lead)

We will run a series of workshops (D2.7) during the evenings of 4 major international meetings on magnetism research[2] to disseminate the micromagnetic virtual research environment (WP4.**T4.9** and

[2]Anticipated most significant international meetings in the appropriate time frame 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. Other training events have been held

T2.5) in the micromagnetic community. Each conference attracts around 1500 participants, and we expect at least 30 for our workshops at every event. Depending on demand, multiple workshops will be given per conference.
The taught material will include (i) use of the Jupyter-based micromagnetic VRE, and an (ii) introduction to the standard techniques for contributing to open source software (version control, pull requests, testing frameworks) to foster excellence in computational science and to make the micromagnetic VRE project self-sustaining as quickly as possible. In addition, all teaching materials, including videos, will be made available on a website.
For each workshop, we request 500 EUR room hire at the magnetism conference location and the travel expenses for two teachers from Southampton to attend the one week international conference, totalling (€500 + 2x€2200=€4900) per workshop. There are no other costs.

T2.7 Workshops in developing countries

In this task, we run a series of workshops in developing countries especially Africa and South America. Some of these workshops will be joined to CIMPA schools. The CIMPA is an international organization based in Nice (France) that promote research in mathematics in developing countries. It organizes each year around 20 schools.
We request €16k to cover expenses for four workshops (each workshop involving two persons of OpenDreamKit).

T2.8 Demonstrator: interactive books Sites: US (lead)

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.**T4.6**. In this task we will demonstrate usability of the results of WP4.**T4.6** 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 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?

T2.9 Demonstrator: Computational mathematics resources indexing service

Beyond official documentation and tutorials, users of mathematical software and VREs learn from a wide array of sources: university courses, Q& A sites, web searches, etc. A simple web search on

in the past at these conferences and were well attended.

any major software component yields dozens of non-official tutorials and how-tos in many different languages. However, search engines mostly miss the relevant metadata: how does one find a tutorial on linear algebra in PARI/GP, written at an undegraduate level, in French or Spanish?

This need has been felt by most communities at some point, and each has come up with its own solution: most software components (e.g., GAP, PARI/GP, Sage, ...) simply link material from their official page; Sage has a wiki (http://wiki.sagemath.org/) referencing additional resources, and used to host a large number of tutorial worksheets on http://sagenb.org/; the recent introduction of public projects in SageMathCloud is sparking approximately the same phenomenon that had previously happened with http://sagenb.org/; Ipython host the Notebook Viewer service (http://nbviewer.ipython.org/), which renders (without hosting) community-made notebooks; teaching institutions host or link their own collections of pedagogical resources. The list goes on.

These collections are usually incomplete, limited in scope, hard to search, outdated, etc. What the community needs is a community-curated, searchable, metadata-driven, multilingual, platform agnostic indexing service whose goal is to reference and rank all the community generated knowledge around a software component or VRE.

The goal of this task is to create the tool (D2.8) powering such service, and to host a (free) community-curated index for OpenDreamKit related resources as a demonstrator (D2.9).

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:

D2.1 (Due: 36, Type: DEM, Dissem.: PU, Lead: ??) *Demo: Nonlinear Processes in Biology interactive book*

D2.2 (Due: 40, Type: DEM, Dissem.: PU, Lead: ??) *Demo: Classical Mechanics interactive book*

D2.3 (Due: 12, Type: DEM, Dissem.: PU, Lead: ??) *Demo: Problems in Physics with Sage v1*

D2.4 (Due: 30, Type: DEM, Dissem.: PU, Lead: ??) *Demo: Problems in Physics with Sage v2*

D2.5 (Due: 32, Type: DEC, Dissem.: PU, Lead: USO) *Micromagnetic VRE code and documents source online*

D2.6 (Due: 44, Type: DEM, Dissem.: PU, Lead: ??) *Demo: Problems in Physics with Sage v3*

D2.7 (Due: 44, Type: OTHER, Dissem.: PU, Lead: USO) *Micromagnetic VRE workshops delivered*

D2.8 (Due: 24, Type: P, Dissem.: PU, Lead: ??) *Community-curated indexing tool (open source)*

[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

D2.9 (Due: 24, Type: DEM, Dissem.: PU, Lead: ??) *Community-curated indexing service for OpenDreamKit*

D2.10 (Due: 18, Type: DEC, Dissem.: PU, Lead: ??) *Short Course: A short course for lecturers on using OpenDreamKit for delivering mathematical education.*

D2.11 (Due: 36, Type: DEM, Dissem.: PU, Lead: ??) *Demo: Interactive lecture notes and marking systems based on OpenDreamKit.*

Work Package 3: Component Architecture						**Start:** 0
Site	UPS	*UVSQ*	USHEF	USO	USTAN	**all**
Effort	64	*8*	4	6	16	**98**

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.

T3.1 Portability Sites: UVSQ (lead)

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 (D3.1, D3.3).
- The deployment of a common infrastructure for multi-platform continuous integration (testing, building and distribution) will be addressed (D3.4).

T3.2 Interfaces between systems Sites: UPS (lead)

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" (`http://www.symbolic-computing.org/`) 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 (D3.5). 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, in D3.6 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.

T3.3 Modularization and packaging Sites: UVSQ (lead)

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. In particular, pushed by cloud computing, containerization and virtualization have become a major trend for distributing complex software, thanks to their ease of installation and configuration. We shall experiment with these technologies by building and distributing virtual appliances for the major components of OpenDreamKit (D3.7).

Once the initial study will have identified the present shortcomings, we will 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. Our first experiment will be to enhance `Sage`'s package system (D3.8), enough to support an open repository of user-contributed code, in the same spirit of modern systems such as Julia (http://pkg.julialang.org/) or PyPI (https://pypi.python.org/). Once *internal* packaging has been dealt with, the route will be paved to further modularize the `Sage` distribution, and make sure that the major Linux distributions have standard packages for it (D3.9).

T3.4 Simulagora integration Sites: Logilab (lead)

Every six month deliver a new Simulagora VM image with as many as possible of the software components released over the period. The goal is to prove that the project is improving the component architecture by measuring the time it takes to integrate them.

T3.5 Component architecture for High Performance Computing and Parallelism

As in all other areas of science, properly supporting massively parallel architecture is a major challenge. Many of the computational components have already gone a long way in this direction, and further work will happen there within WorkPackage 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. We will use `Sage` and its components as a test-bed, by producing an HPC-enabled distribution (D3.10).

T3.6 Document and modularize `SageMathCloud`'s codebase Sites: UPS (lead)

From its inception in 2013, `SageMathCloud` has quickly developed to a full featured VRE. Because of the tight, partly closed source development cycles, `SageMathCloud`'s codebase has quickly grown in size, with its documentation not always keeping the pace. As a result, it is at the moment very hard for a newcomer to setup a clone service of `SageMathCloud` just from its sources.

Now that `SageMathCloud` is fully open source, we need to go through its codebase, understand and document it (D3.11), isolate components that might be reused by other software (e.g.: `Jupyter`), and make it as portable as possible.

The ultimate goal of this task is to produce a *personal* version of `SageMathCloud`, to be shipped along with `Sage`, that a user can run on his own personal computer (D3.12).

T3.7 Improving the development workflow in mathematical software Sites: UVSQ (lead)

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`: D3.13).).

T3.8 Python interface for OOMMF micromagnetic simulation library Sites: USO (lead)

In this task, we create a Python interface (D3.2) for the open source Object Oriented MicroMagnetic Framework (OOMMF [1]). As a result, the OOMMF library will be fully accessible and usable from a Python interface and become a component in the Python/`Jupyter` eco system of computational tools and in OpenDreamKit. We make use of this component architecture in WP4.**T4.8**.

In more detail, we will first identify the 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 C++ objects in Python, providing an architecture component that provides full access to OOMMF's raw capabilities. For clarity, we will refer to this interface as `OOMMF-py-raw`.

Secondly, we will create a user friendly Python library `OOMMF-py` that combines the `OOMMF-py-raw` capabilities in an object orientated and safe-to-use Python library targetting researchers in the magnetic materials community. We will follow the design of the well-received high level Python interface in the Nmag micromagnetic simulation package [8] interface [21]. Unit and regression tests for both component interfaces `OOMMF-py-raw` and `OOMMF-py` are simultanously developed.

Deliverables:

D3.1 (Due: 6, Type: OTHER, Dissem.: PU, Lead: ??) *One-click install Sage distribution for Windows with Cygwin 32bits*

D3.2 (Due: 9, Type: OTHER, Dissem.: PU, Lead: USO) *Python Interface to micromagnetic OOMMF package completed*

D3.3 (Due: 24, Type: OTHER, Dissem.: PU, Lead: UPS) *One-click install Sage distribution for Windows with Cygwin 64bits*

D3.4 (Due: 36, Type: DEM, Dissem.: PP, Lead: ??) *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.

D3.5 (Due: 12, Type: OTHER, Dissem.: PU, Lead: USTAN) *Support for the SCSCP interface protocol in all relevant components (Sage, GAP, etc.) distribution*

D3.6 (Due: 36, Type: OTHER, Dissem.: PU, Lead: UO) *Semantic-aware Sage interface to GAP.*

D3.7 (Due: 6, Type: OTHER, Dissem.: PU, Lead: ??) *Virtual images and containers* Creation and distribution of preconfigured cloud oriented virtual machines/containers (e.g. Docker images) for PARI/GP, Sage, SageMathCloud, ...

D3.8 (Due: 24, Type: OTHER, Dissem.: PU, Lead: ??) *Open package repository for Sage* Refactor Sage package and build system to support community-contributed packages installable via the web.

D3.9 (Due: 48, Type: OTHER, Dissem.: PU, Lead: ??) *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.

D3.10 (Due: 48, Type: OTHER, Dissem.: PU, Lead: ??) *HPC enabled Sage distribution*

D3.11 (Due: 6, Type: R, Dissem.: PU, Lead: ??) *Understand and document SageMathCloud backend code.*

D3.12 (Due: 24, Type: OTHER, Dissem.: PU, Lead: ??) Personal *SageMathCloud: single user version of SageMathCloud distributed with Sage.*

D3.13 (Due: 24, Type: OTHER, Dissem.: PU, Lead: ??) *Integration between SageMathCloud and Sage's TRAC server*

Work Package 4: User Interfaces										**Start:** 0
Site	UPS	Logilab	UVSQ	CNRS	USHEF	USO	USTAN	JacU	UK	*Simula*	**all**
Effort	26	1	2	1	12	16	18	22	1	*23*	**122**

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.

The `Jupyter` notebook is being used in all areas of academic (including for example University of California, Berkeley, Stanford, MIT, Harvard, Cambridge, Oxford, Imperial College, Southampton, Hamburg, Paderborn, Vienna, Paris, Katowice, and Oslo) 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, theses) 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.

To demonstrate the power of the OpenDreamKit environment to accelerate computational science, deliver better value for money and make computational science more robust, we will put together a micromagnetic Virtual Research Environment (micromagnetic VRE) 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] as there is a wide community to benefit from it.

T4.1 Uniform notebook interface for all interactive components Sites: UPS (lead)

In this task, we will implement `Jupyter` interfaces for the interactive computation components of OpenDreamKit, including GAP, Pari, `Sage`, and Singular. A first release D4.4 will focus on basic functionality, and a second release D4.5 will cover advanced features like 3D graphics or transparent documentation browsing (as live worksheets whenever relevant).

`Sage` itself will require a specific treatment as it already has a notebook interface. Its development started about at the same time as the `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 D4.6 will require:
- Robust migration path and tools for Sage worksheets,
- Support for math, 2D, and interactive 3D scene visualization,
- 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.

T4.2 Notebook improvements for collaboration Sites: Simula (lead)

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 D4.11.
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 D4.17.

T4.3 Reproducible Notebooks Sites: Simula (lead)

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 D4.10. Prior work has been done in Sage for ReST files, e.g. sage −t notebook.rst, and this model will be extended to notebooks.

T4.4 Refactor Sage's Sphinx documentation system Sites: UPS (lead)

Sage, like Python and many other Python based projects, uses the Sphinx documentation system. Due to particularly stringent needs, many layers of customization and adaptations have accumulated over the years, in particular for proper scaling to the sheer size of the Sage documentation (13k pages just for the reference manual).
A deep refactorization (D4.18) is critically needed to get rid of multiple duplication, and foster sustainability by outsourcing back to Sphinx all generic aspects (parallel compilation, index generation, ...).

T4.5 Dynamic documentation and exploration system Sites: UPS (lead)

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.
- Widgets based on the HTML5 and web component standards to display graphical views of the results of SPARQL queries, as well as populating data structures with the results of such queries,
- D4.19 (Month 36) Exploratory support for semantic-aware interactive widgets providing views on objects of the underlying computational or database components. Preliminary steps are demonstrated in the Larch Environment project (see demo video on http://www.larchenvironment.com/) and sage-explorer (https://github.com/jbandlow/sage-explorer). The ultimate aim would be to automatically generate LMFDB-style interfaces.

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

T4.6 Structured documents Sites: JacU (lead)

Jupyter notebook is basically a sequence of cells that contain either text or a program. The complex documents, such as books, articles or reports, require a richer description that covers the the structure of the document and the semantics of its elements. This task will investigate this problem and try to find a way to write these documents using the breakthroughs achieved in the other tasks to this workpackage.

Several technical complementary options can be explored:

- 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,. . .)
- Jupyter notebooks are essentially “programs with documentation’ and lack the semantical structure needed by complex documents.
- sTeX is a semantic variant of LaTeX that can be transformed into OMDoc/MMT, which is the native knowledge representation format for active documents and machine-actionable knowledge about math and symbolic programs.

After gathering the needs and the requirements for the writing of complex documents in the mathematical field, we will study these various options, design and build a solution that meets the expectations. This work will be done through an iterative process that tries to enhance little-by-little the already existing softwares.

T4.7 Active Documents Portal Sites: JacU (lead)

We will extend the existing http://mathhub.info systeml to a portal for interacting with active/structured documents (see **T4.6**) and run it for internal use in the OpenDreamKit first and later for general use outside the project. MathHub.info alread provides very basic sTeX editing and verisoning. For OpenDreamKit we will need to extend it with the computational side based on the integrated format from **T4.6**.

T4.8 Case study: micromagnetic VRE built from OpenDreamKit Sites: USO (lead)

In this task, we use the OpenDreamKit architecture to assemble a virtual research environment software tailored for the large community of researchers in magnetic materials that use simulation studies for their work (see introduction in description on page 45). These researchers in academia and industry are typically not computationalists and use simulation to support their experiments and analytical work.

The micromagnetic VRE will be based on the Jupyter Notebook, the Python interface to the micromagnetic simulation library OOMMF (WP3.**T3.8**), and the additional features added to Jupyter in this work package. The pool of several thousand OOMMF users are the direct beneficiaries of this demonstrator, but we will extract further and more generic lessons and insight from it (for example WP7.**T7.4**).

The Jupyter Notebook environment allows to host, execute and document the Python-based OOMMF simulation in an executable document. In this interactive environment, objects can be displayed using various representations, including, for example, textual representation (i.e. strings), bitmap images and SVG (vector graphics) files. We will create 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 (linking to **T4.11**). This allows computational steering and the interactive exploration of the behavior of magnetic nanostructures.

Beyond that, the Jupyter Widgets allow the creation of graphical user interface (GUI) elements, and we will generate code to display these widgets on demand to (i) set up micromagnetic simulations using a GUI, and (ii) assist in common post-processing simulation results. 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 [22] 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, caters for users' preferences, and provides significant additional value.

The deliverable for this task is the open source micromagnetic VRE software (D4.7).

T4.9 Demonstrator: micromagnetic VRE notebooks Sites: USO (lead)

The purpose of the micromagnetic VRE (**T4.8**) is to enable excellent computational research. To maximise the value of this grant's investment for the community, we will not carry out micromagnetic research but instead produce a set of executable notebooks using the micromagnetic VRE to demonstrate its power and applicability.

We will create executable notebook documents (D4.12) within the micromagnetic VRE including (i) a new tutorial on usage of OOMMF, (ii) the complete documentation of the OOMMF-Py library, and (iii) a set of typical micromagnetic case studies. The tutorial, in terms of content, will take guidance from the tutorial provided for Nmag [20] but is tailored for the special simulation capabilities of underlying micromagnetic OOMMF library, and will introduce the additional features of the Jupyter-driven micromagnetic VRE. We expect this substantial and executable documentation of the micromagnetic VRE to become the standard resource that introduces researchers to computational micromagnetics, in particular through the online portal (**T4.10**).

T4.10 Online portal for micromagnetic VRE demonstrator Sites: USO (lead)

Recently, a TeMPorary Jupyter NoteBook (TMPNB) 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 such a portal (D4.14) which provides the micromagnetic VRE for anonymous use. This service allows users to execute the demonstrator tutorial and documentation notebooks (**T4.9**) and run the calculations in real time on the web server, without having to install any software on their own machine. This webservice will be based on Docker [6] virtualisation technology and we will make available the scripts to create VirtualBox [29] images, and Docker containers. The same virtual machine images can also be used for Cloudhosted computing services.

We request 6000 EUR to purchase a machine to provide these services (shared memory, 64 cores, 128GB RAM, Solid-state drive (SDD) to make the system more responsive).

T4.11 Visualization system for 3d data in web-notebook

The Jupyter notebook is a very attractive environment for building user interfaces for research, but the current support for inline visualization is limited to curve plots and 2d scalar fields. Many scientific simulations need visualization of 3d scalar and vector fields. The amount of data can be tremendous, especially in time-dependent problems computed in a distributed fashion over large-scale computational clusters. Interactive inspection of such simulations can be a valuable tool which accelerates research. For inspection, one does not need to transfer and gather the full dataset at each time step - getting selected computed fields on user request or preprocessing certain quantities like cross sections with some predefined frequency will mostly suffice.

We propose to let computational fluid dynamics (CFD) be a driving application for the development of 3d visualization in Jupyter notebooks since CFD is one of the most demanding cases of scientific visualization. Successfully handling CFD makes the tool immediately applicable to a range of other fields such as heat transfer, electromagnetics, material science, and 3d algebraic structures in mathematics. Simulations are started inside the notebook and executed on HPC clusters. This approach will significantly lower the threshold for using parallel computing codes that can be hard to install correctly on local workstations. Such use cases with 3d visualization will greatly extend the potential applications of the `Jupyter` notebook concept throughout science and engineering.

As example code for a 3d live web notebook with fluid dynamics simulations, we will use the Lattice Boltzmann solver which is under development at the University of Silesia: Sailfish. This code is an advanced Lattice Boltzmann solver designed from the ground up for distributed system of GPU compute clusters. It is implemented predominantly in Python, and 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.

The project must first investigate available technologies for fast in-browser visualization via Vispy and then determine the most appropriate visualization techniques for the typical structures to be displayed (isosurfaces, streamlines, flow sheets, etc.). After a proof of concept is established by researchers, a precise description of the visualization tool can be subcontracted to professional IT developers.

T4.12 Math Search Engine Sites: JacU (lead)

T4.13 Simulagora collaboration Sites: Logilab (lead)

Comparing the collaboration workflow of Simulagora, SageMathCloud and GitHub.

Deliverables:

D4.1 (Due: 6, Type: R, Dissem.: PU, Lead: JacU) *Active Documents based on sTeX*

D4.2 (Due: 9, Type: OTHER, Dissem.: PU, Lead: JacU) *Full-text search (formulae + Keywords) over LaTeX-based documents (e.g. the arXiv subset)*

D4.3 (Due: 12, Type: DEM, Dissem.: PU, Lead: JacU) *Distributed, Collaborative, Versioned Editing of Active Documents in MathHub.info*

D4.4 (Due: 12, Type: OTHER, Dissem.: PU, Lead: ??) *Basic Jupyter interface for GAP, Pari, Sage, Singular*

D4.5 (Due: 12, Type: OTHER, Dissem.: PU, Lead: ??) *Full featured Jupyter interface for GAP, Pari, Singular*

D4.6 (Due: 12, Type: DEM, Dissem.: PU, Lead: ??) *Sage notebook / Jupyter notebook convergence*

D4.7 (Due: 15, Type: DEM, Dissem.: PU, Lead: USO) *Micromagnetic VRE software completed*

D4.8 (Due: 18, Type: OTHER, Dissem.: PU, Lead: JacU) *Full-text search (Formulae + Keywords) over Notebooks* We want to be able to search over the Jupyter-based Notebooks from **T4.6**.

D4.9 (Due: 18, Type: DEM, Dissem.: PU, Lead: JacU) *In-place computation in active documents (context/computation)*

D4.10 (Due: 18, Type: OTHER, Dissem.: PU, Lead: Simula) *Facilities for running notebooks as verification tests*

D4.11 (Due: 12, Type: OTHER, Dissem.: PU, Lead: Simula) *Tools for collaborating on notebooks via version-control*

D4.12 (Due: 21, Type: DEC, Dissem.: PU, Lead: USO) *Micromagnetic VRE tutorial and documentation notebooks*

D4.13 (Due: 24, Type: DEM, Dissem.: PU, Lead: JacU) *Notebook Import into MathHub.info (interactive display)*

D4.14 (Due: 24, Type: DEC, Dissem.: PU, Lead: USO) *Demonstrator online portal available*

D4.15 (Due: 24, Type: OTHER, Dissem.: PU, Lead: ??) *Jupyter extension for 3d CFD visualization*

D4.16 (Due: 30, Type: OTHER, Dissem.: PU, Lead: JacU) *Formula search in CAS programs and Software Modules*

D4.17 (Due: 36, Type: OTHER, Dissem.: PU, Lead: Simula) *Exploratory support for live notebook collaboration*

D4.18 (Due: 24, Type: OTHER, Dissem.: PU, Lead: ??) *Refactorization of Sage's Sphinx documentation system*

D4.19 (Due: 36, Type: DEM, Dissem.: PU, Lead: ??) *Exploratory support for semantic-aware interactive widgets providing views on objects represented and or in databases*

D4.20 (Due: 42, Type: OTHER, Dissem.: PU, Lead: JacU) *Search from Notebooks/Active Documents Interface* Often it is important to have some local context to inform search, therefore search from the notebook/active documents interface is an interesting and useful development target.

Work Package 5: High Performance Mathematical Computing							**Start:** 0
Site	UPS	Logilab	*UJF*	CNRS	USTAN	UK	**all**
Effort	6	12	*52*	1	18	1	**90**

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 architectures is a major challenge. Many of the computational components in OpenDreamKit have already gone a long way in this direction. For example, parallel versions of the GAP kernel for a range of architectures were developed during the 2009-2013 EPSRC "HPCGAP" 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 HPC-GAP. 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 component Tasks **T5.1**,**T5.2**, **T5.3**, **T5.5** and **T5.4**.

Many of the computational components of OpenDreamKit, notably Sage and GAP use a high level interpreted language for their library. Performance is achieved by rewriting or compiling critical sections into a lower-level language. Sage uses the Cython Python-to-C compiler; GAP has some more basic support. In Tasks **T5.2**, ?? and ??, we will also boost performance by further developing and applying such compilation tools, allowing the application programmer to retain their high level approach.

T5.1 PARI Sites: CNRS (lead)

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.

T5.2 GAP Sites: USTAN (lead)

Thanks to the HPCGAP project, almost the full functionality of GAP can be safely run on multicore architectures, and there is support for simple but effective parallel programming, with protection from most of the more serious pitfalls that can trouble the novice parallel programmer. Experimental versions of GAP also exist for a number of distributed-memory architectures.

In this task we will continue to develop the GAP infrastructure to offer performance improvements to real end users on a wide range of modern hardware. This will be achieved by a number of synergistic developments in the system

- a library of parallel algorithms for algebraic computation. This will include general purpose skeletons applicable to many problems; distributed data structures suitable for orchestrating distributed memory computations; and implementations of specific parallel algorithms for key mathematical tasks.

- interfaces between GAP and standard cloud and HPC infrastructure. This work will be based on WP3.**T3.5**.
- adaptation of the cython and/or pythran technology to allow the performance of critical GAP language subroutines to be increased close to that of C code without the programmer cost of a full C implementation.

T5.3 Linbox Sites: UJF (lead)

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.

T5.4 Singular Sites: UK (lead)

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.

T5.5 MPIR Sites: UK (lead)

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.

T5.6 HPC infrastructure for combinatorics Sites: UPS (lead)

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 most important feature is that the goal of research is mostly to design and to understand properties of algorithms. As a consequence, much more often that in other field, the researcher needs to program. However, this is not his ultimate goal so the programming environment must be very expressive for fast prototyping.

At the same time, the problems often require relatively large computations; algorithms of exponential complexity are extremely common, and combinatorial explosion is the main obstacle to many experiments. Hence the programming environment must make no compromise on efficiency.

Finally, embarrassingly parallel problem are extremely common, and more often than not problems that are not embarrassingly parallel reduce to the exploration of a large tree. Hence the programming environment must minimize the extra work needed to get from a serial program to a parallel one in these simple situations.

Through this task we will provide a concrete, practical, and highly demanding use cases for the infrastructure developed in this work package. In particular, they will serve to evaluate the benefits of tasks ??, WP3.**T3.3**, and WP3.**T3.5**.

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++`.

T5.7 Pythran Sites: UJF (lead)

`Cython` is a extended-`Python` to C compiler that was originally developed for `Sage` and is now a thriving project of its own.

`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.
- The optimized runtime of `Pythran` can be used through `Cython`.

An effort will be made to improve more and more the parallelism in the `Pythran` runtime D5.4.

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.

Adding `Pythran` support in `Sage` will be done not only for `Sage` code but also for `Sage` users code thanks to compilation facilities in the notebook interface. D5.1

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 to benefit `Pythran` compile time optimisation. A specific challenge is that the `Sage` library uses quite heavily object-oriented programming.

This task will strongly benefit from Task ??, 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 D5.6.

T5.8 Sun Grid Engine Integration in Project Jupyter Hub Sites: USHEF (lead)

The Sun Grid Engine is a commonly used High Performance Computing scheduler. It is used, for example, on the institutional HPC systems in both Sheffield and Manchester Universities as well as the regional N8 HPC facility, a system shared by the 8 most research intensive universities in the North of England. In this task, we will develop and demonstrate a Sun Grid Engine notebook spawner for Project Jupyter, allowing users to access Jupyter notebooks on the HPC cluster. This will enable the interactive analysis of output data products on the cluster where they were generated and are stored, via a user-friendly web interface.

Deliverables:

D5.1 (Due: 3, Type: DEM, Dissem.: PU, Lead: ??) *Facility to compile Pythran compliant user kernels and sage code.*

D5.2 (Due: 3, Type: DEM, Dissem.: PU, Lead: ??) *Parallelise the relation sieving component of the Quadratic Sieve*

D5.3 (Due: 3, Type: DEM, Dissem.: PU, Lead: ??) *Turn the Python prototypes for tree exploration into production code, integrate to Sage.*

D5.4 (Due: 9, Type: DEM, Dissem.: PU, Lead: ??) *Improve Pythran runtime support to automatically take advantage of multi-cores and SIMD instruction units and use it in Cython.*

D5.5 (Due: 9, Type: DEM, Dissem.: PU, Lead: ??) *Implement a parallel version of Block-Wiederman linear algebra over GF2 and the triple large prime variant.*

D5.6 (Due: 12, Type: DEM, Dissem.: PU, Lead: ??) *Make Pythran typing better to improve error information.*

D5.7 (Due: 12, Type: DEM, Dissem.: PU, Lead: ??) *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*

D5.8 (Due: 12, Type: DEM, Dissem.: PU, Lead: ??) *Parallelise the Singular sparse polynomial multiplication algorithms*

D5.9 (Due: 12, Type: DEM, Dissem.: PU, Lead: ??) *Parallel versions of the Singular sparse polynomial division and GCD algorithms.*

D5.10 (Due: 12, Type: DEM, Dissem.: PU, Lead: ??) *Extend the existing assembly superoptimiser for AVX and upcoming Intel processor extensions*

D5.11 (Due: 12, Type: DEM, Dissem.: PU, Lead: ??) *Provide FAT binary support for all new x86_64 processors, build system maintenance and improvements to tuning, profiling and testing subsystems*

D5.12 (Due: 12, Type: DEM, Dissem.: PU, Lead: ??) *Explore the possibility to interface smoothly Pythran, Cython and Cilk++.*

D5.13 (Due: 12, Type: R, Dissem.: PU, Lead: UJF) *Library design and domain specific language exposing LinBox parallel features to Sage*

D5.14 (Due: 18, Type: R, Dissem.: PU, Lead: USTAN) *Report on development of designs for the GAP developments – parallel library, interacts to standard infrastructure and Cython-like extensions*

D5.15 (Due: 24, Type: DEM, Dissem.: PU, Lead: ??) *Ongoing support of Intel, AMD, ARM, Sparc processors and new Operating System versions*

D5.16 (Due: 24, Type: DEM, Dissem.: PU, Lead: ??) *Refactor the existing combinatorics Sage code using the new developed Pythran and Cython features.*

D5.17 (Due: 24, Type: DEM, Dissem.: PU, Lead: UJF) *Algorithms and implementations. Library maintenance and close integration in mathematical software*

D5.18 (Due: 36, Type: DEM, Dissem.: PU, Lead: UJF) *Implementations of algorithms using distributed memory et heterogenous architectures: clusters and accelerators*

D5.19 (Due: 47, Type: OTHER, Dissem.: PU, Lead: USTAN) *Implementations of the GAP developments, ready for release*

D5.20 (Due: 48, Type: R, Dissem.: PU, Lead: USTAN) *final report and evaluation of the GAP devel-*

opments

Work Package 6: Data/Knowledge/Software-Bases								Start: 1
Site	UPS	Logilab	USHEF	USTAN	UW	*JacU*	UZH	**all**
Effort	4	2	12	10	25	*36*	12	**101**

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, traceability, versioning and visualisation.

Description

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 WP4, 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.

T6.1 Survey of existing databases Sites: UZH (lead), JacU, USTAN, UW, US

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.

T6.2 Formulation of requirements and design of new infrastructure when appropriate Sites: JacU (lead), UZH, US, USTAN, UW, Logilab

Ontologies are the canonical method used to implement databases that require significant data interchange. However, because of the extreme reification present in mathematics (relations between objects themselves become objects of study), there are specific obstacles compared to the usual semantic web model of publishing.

Drawing on semantic web/Linked Open Data experience of the Logilab group, specialised to mathematics through the OMDoc/MMT work of the Bremen group, we will design a decentralised infrastructure for OpenDreamKit. This infrastructure would allow modular collaborations, through decentralised hosting of data without the need to merge everything centrally.

We will organise a workshop associated to this task.

T6.3 Triform Theories in OMDoc/MMT Sites: JacU (lead), UZH
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.

T6.4 Computational Foundation for Python/Sage (or some CAS) Sites: JacU (lead), UZH, USTAN
In the OMDoc/MMT world a foundation is a logical base language that gives the formal meaning to all objects represented/formalized in it. We have created a very initial computational foundation for Scala and implemented it in the MMT API. This can be used to execute (or verify) computations directly in OMDoc/MMT and thus forms the basis for various integration tasks for OMDoc/MMT biform theories that integrate Scala computations. Here we propose to develop a somewhat more complete computational foundation for Python and/or parts of Sage (coverage to be determined). Bi/Triform theories come in three parts:

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

T6.5 OEIS Case Study (Coverage and automated import) Sites: JacU (lead)
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 (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.

T6.6 FindStat Case Study (triformal theories) Sites: JacU (lead), UZH
In this task we would develop triformal theories for the FindStat project to test the design from **T6.4**. 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 , this will again allow for easier knowledge management services, and in particular improved search services.
This Task will be co-developed with **T6.4**, it will validate the design of triformal theories and be iterated to test the design changes.

T6.7 LMFDB Case study (triformal theories) Sites: JacU (lead), UZH
In this task we would develop triformal theories for an exemplary part of the LMFDB project to test the design from **T6.4**. 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 **T6.4**. (decrease importance of verification as opposed to interoperability)

T6.8 Memoization and production of new data Sites: USTAN (lead), US, UPS
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:

D6.1 (Due: 9, Type: R, Dissem.: PU, Lead: JacU) *Workshop Report*

D6.2 (Due: 9, Type: OTHER, Dissem.: PU, Lead: JacU) *Heuristic Parser for the OEIS*

D6.3 (Due: 12, Type: R, Dissem.: PU, Lead: JacU) *Design of Triform (DKS) Theories (Specification/RNC Schema/Examples)*

D6.4 (Due: 12, Type: DEC, Dissem.: PU, Lead: JacU) *Python/Sage Syntax Foundation Module in OMDoc/MMT*

D6.5 (Due: 12, Type: DEC, Dissem.: PU, Lead: UZH) *Conversion of existing and new databases to unified interoperable system*

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

D6.6 (Due: 12, Type: R, Dissem.: PU, Lead: UZH) *LMFDB deep modelling: Fragment Identification & Initial Model Design*

D6.7 (Due: 18, Type: R, Dissem.: PU, Lead: UZH) *LMFDB Data vs. Knowledge vs. Software Validation*

D6.8 (Due: 18, Type: R, Dissem.: PU, Lead: JacU) *Cross-Validation for OEIS DKS-Theories*

D6.9 (Due: 24, Type: OTHER, Dissem.: PU, Lead: UZH) *Shared persistent memoization library for Python/Sage* Recomputation?, Ease of publishing, importing, ...

D6.10 (Due: 24, Type: OTHER, Dissem.: PU, Lead: JacU) *Implementation of Triform Theories in the MMT API.*

D6.11 (Due: 24, Type: OTHER, Dissem.: PU, Lead: JacU) *Python/Sage Computational Foundation Module in OMDoc/MMT*

D6.12 (Due: 36, Type: OTHER, Dissem.: PU, Lead: JacU) *Python/Sage Declarative Semantics in OMDoc/MMT*

D6.13 (Due: 36, Type: OTHER, Dissem.: PU, Lead: JacU) *LMFDB Algorithm verification wrt. a Triformal theory*

D6.14 (Due: 46, Type: R, Dissem.: PU, Lead: JacU) *LMFDB full integration of algorithms, data and presentation (not so much verification)*

database access to LMFDB as a python library

Work Package 7: Social Aspects				**Start:** 0
Site	*UO*	USHEF	USO	**all**
Effort	*53*	8	6	**67**

Objectives

The processes by which mathematical knowledge and mathematical software are developed, validated and applied are quite distinctive. In other sciences, the universe provides "ground truth" and the scientific texts or theories can be validated against that by experiment. In mathematics the text is the ground truth. Once a mathematician, or a small group of mathematicians standing around a blackboard, produced a proof they would "clean it up" removing all traces of the process by which it had been discovered and then submit the "clean" text to their peers for review.

Mathematicians have adopted new technology is a range of ways: email and shared documents are used to collaborate on problem-solving and writing; larger "crowdsourcing" [25, 27], arrangements pull together diverse experts; symbolic computation tackles huge routine calculations; and computers check proofs that are just too long and complicated for any human to comprehend. This technology is both revealing (since email messages, version control systems and bulletin boards can be analyzed) and altering the ways in which mathematicians collaborate.

In an EPSRC funded project "The Social Machine of Mathematics" Martin and others are bringing rigorous methods from the social sciences together to study these processes. This workpackage has the following objectives:

- to bring the insights of this and similar projects into the design of OpenDreamKit VRE, ensuring that it supports the ways in which mathematicians really work, and not just the way software developers or indeed mathematicians, think they do;
- to extend this work to study the collaborative processes of free (mathematical) software development to produce guidelines for best practice and ideas for how existing processes can extend to a "system of systems".
-

Mathematics provides a common language amongst the quantitative sciences, however presentations of mathematics, and mathematical models, are not always consummate with clear propagation of ideas that the mathematics encode. By closely intertwining the application with the mathematics, the project will develop more widespread understanding of the utility of a mathematical abstraction and encourage repeatability of analysis. OpenDreamKit should enable user-friendly presentation of the abstraction to such an extent that even non computational/mathematical scientists feel compelled to interact with and explore the model.

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 [18] and in Polymath Projects [25, 27]. "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 [19] 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 — Sage. We will apply preference and opinion aggregation techniques [4] to develop a community prioritisation scheme for Sage bugs and features requests, which are presently being maintained on the Sage TRAC server [3] and implement them as a TRAC [28] 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 [7], Open source systems are only marginally better in this way, as recent computer security scares, such as the one around Bash [5], indicate. A game-theoretic analysis of this situation will be attempted.

T7.1 Modern Distribution of Scientific Output

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. We will develop live posters for distribution of knowledge, designed for integration with either large touch screens or smaller tablets D7.2. We will augment the Jupyter project with facilities for providing comment on notebooks to encourage debate on mathematical and computational ideas ??.

T7.2 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, Sage.

T7.3 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).

T7.4 Evaluation of Micromagnetic VRE Sites: USO (lead)

We will use the micromagnetic VRE demonstrator (WP4.**T4.9**), its dissemination workshops (WP2.**T2.6**) and interactions with its users and contributors in the micromagnetic community to evaluate, reflect and report on the project, taking into account technical and social aspects.

A survey will be developed and used to gather user input and feedback on usefulness of the provided capabilities, with particular focus on the capabilities of the micromagnetic VRE to (i) enable new and better science, to (ii) allow to make progress effectively, to (iii) carry out computational science reproducibly, to (iv) collaboratively enable trust and to (v) become a self-sustained project from community contributions. Amongst other channels, we will target attendees of the micromagnetic VRE dissemination workshops (WP2.**T2.6**) to gather data.

All results and insights will be summarised in a public document (D7.3) and reported at appropriate workshops and conferences to share the lessons learned from this Jupyter-based VRE for micromagnetics. We will create a manuscript for journal publication, summarising the demonstrator project and this evaluation. An important point of this publication is to provide a reference that can be cited by publications making use of the new micromagnetic VRE, to allow tracking of uptake and development of this VRE beyond the life time of this H2020 project.

Deliverables:

D7.1 (Due: 36, Type: DEM, Dissem.: PU, Lead: USHEF) *Demonstrator: Jupyter Notebook Live Poster*

D7.2 (Due: 24, Type: DEM, Dissem.: PU, Lead: USHEF) *Demonstrator: Mechanism for comment on posted Jupyter notebooks.*

D7.3 (Due: 48, Type: R, Dissem.: PU, Lead: USO) *Micromagnetic VRE environment evaluation report*

3.3 Management Structure and Procedures

3.3.1 Management

The project will be coordinated by the University Paris-Sud (UPSud), represented by Prof. Nicolas Thiery (Project Coordinator), who has experience in successfully managing several research projects on the main OpenDreamKit topics. Pioneer in community-developed open source software for research in this field, Thiery founded in 2000 the Sage-Combinat software project involving 50 researchers in Europe and abroad. This project has grown under his leadership to be one of the largest organized communities of Sage developers.

The Project Coordinator will be assisted by a part-time (50%) Project Manager, that will be hired for this project and located in the European Affairs and Technology Transfer Office (SAIC) of the UPSud. Additional feedback and expertise will be brought by Financial, Legal and European affairs officers from SAIC.

3.3.2 Organizational structure and decision-making

The organizational structure, shown in the Figure XXXXXXX, has been designed to enable efficient coordination of the OpenDreamKit project — a VRE integrating several previously separated tools and software and involving both academic actors and industrial stakeholders.

We have designed the management structure and procedures to deal in a flexible manner with the following challenges:

- to integrate all consortium members and to mobilize their expertise, knowledge and networks at every stage of the project;
- to give the maximum attention to the end-users needs and requirements;
- to continuously involve expertise and knowledge of relevant stakeholders and their networks, and
- to efficiently coordinate the project implementation in a collaborative environment and ensure its sustainability.

The coordinator is acting as an intermediary between the Partners and the European Commission. The coordinator will oversee the project planning, monitor that the execution is carried out in time and that the objectives are achieved and closely interact with the project officer for the project monitoring and delivering the performance indicators. The Project Manager will ensure an efficient day-to-day management of the project, reporting, feedback to partners on administrative, financial and legal issues, follow-up of the resources allocation and consumption, communication inside and outside the consortium.

The resources of all partners will be mobilized by decentralisation of responsibilities through the assignment of leadership for work packages. A clear task sharing obtains efficient decision making mechanisms and a sound financial management will safeguard the achievement of the project's objectives.

3.3.3 Project roles

The following bodies will form the organizational structure of the OpenDreamKit project : Coordination Team (MT), Steering Committee (SC), Advisory Board (AB), End User Group (EUG) and Quality Review board (QRB).

Coordination Team (CT)

Members: The CT is composed of the Work Package leaders and headed by the Project Coordinator, assisted by the Project Manager.

Responsibilities: CT is an executive body in charge of the project implementation and monitoring. It takes operational decisions necessary for the smooth execution of the project.
Tasks:

1. Monitoring the timely execution of the tasks and achievement of the objectives;
2. Preparing of scientific and financial progress reports;
3. Controlling the WP progress by assessing it through technical reports developed by the WP partners;
4. Making proposals to the Steering Committee of re-allocation of tasks, resources and financial needs for the fulfilment of the work plan;
5. Preparing the drafts and validating the project deliverables to be submitted to the Commission; Meetings : Project Coordinator and Project Manager can meet any time and at least twice a week. They will meet Work-Package leaders every 6 months. If necessary, extra-meetings will be arranged.

Steering Committee (SC)

Members: The SC is chaired by the Project Coordinator and includes one representative from each partner organization.

Responsibilities: The SC is the decision-making body in charge of its strategic orientation. It takes decisions on scientific orientations of the project, re-allocation of resources, consortium changes, intellectual property rights

Meetings: Every 6 months. If necessary, extra-meetings will be arranged. Written minutes of each meeting will be produced, which shall be the formal record of all decisions taken. A procedure for comment and acceptance is proposed.

Voting procedure: The SC shall not deliberate until all Members are present or represented. Each Member shall have one vote. The GA will work on consensual decisions as much as possible and resort to voting only if unavoidable. Voting decisions shall be taken by a majority of two-thirds (2/3) of votes.

Advisory board (AB):

Members: top level experts from partner organizations and external, including both experts from the project scientific area, and experts on legal and social matters.

Responsibilities: to give an independent opinion on scientific and innovation matters, in order to guaranty quality implementation of the project, efficient innovation management and project sustainability. Meetings : on the demand of the Steering Committee.

Quality Review Board (QRB)

Members: The QRB will be composed of 2 senior researchers from the Consortium, 2 representatives of the End User Group and 2 experts from the Advisory Board. It will be chaired by one of the professors within the Consortium. All members will be appointed at kick-off meeting of the project.

Responsibilities: to monitor the quality of the Deliverables, the whole ‘production process’ and to recommend improvements during the project to the SC.

Meetings: before publications and reports of the project.

End User Group (EUG)

Members: end-users of the VRE, internal and external to the consortium, from different disciplines and both from academic and industrial sector. They are actively involved into the project execution, and work in close interaction with the project coordinator.

Responsibilities: the EUG is the main actor of the innovation management within the consortium, as they have a deep understanding of both market and technical problems, and awareness of opportunities. The EUG also plays a main role in ensuring the VRE sustainability.

Tasks: to control the project execution from the point of view of the end user needs and requirements, to test the tool and to detect its potential shortcomings at the early stages, to propose adaptation measures. Meetings : the EUG will have regular virtual meetings, and will meet physically at least once a year.

3.3.4 Project management tools and procedures

Project partners and management bodies will communicate through especially dedicated project web platform, maintained by the Project Manager. WP leaders will at least monthly monitor progress of participants of their WP, and participants will inform their WP leaders when problems are encountered. Major problems will be discussed in (teleconference) meetings with the project Coordinator and Project Manager. Each WP leader will be free to organize extra-meetings with WP partners, if necessary. Scientific and financial progress reports will be collected, assembled and transmitted to the Project coordinator by the WP leaders through the web platform. On basis of the Progress Reports, the Coordination Team will monitor progress of the project, identify bottlenecks and find solutions for these problems. Where needed, adaptations to the project plan will be made, with the aim to ensure the delivery of the project results as agreed with the EC. Major adaptations need to be approved by the Steering Committee. If necessary, the SC can submit reports to the QRB for opinion.

Finally, the End Users Group, working in close cooperation with the project coordinator, will ensure the efficient innovation management. They will carefully monitor the new opportunities, in order to give, if necessary, new orientations to the project. For legal aspects, they will have a feedback from legal officers from the Coordinator's European Affairs and Technology Transfer office (SAIC), specialized in Intellectual Property.

Our management structure and procedures will ensure that our network of 16 partners from both academic and industrial sectors is focused at achieving the promised deliverables, efficiently managing the innovation process and largely opening the VRE to its final users. The 15 EU-partners will sign a Consortium Agreement, in which operational rules and decision making procedures will be laid down. The international partner will work with a bi-lateral agreement with the Coordinator.

3.3.5 Risk management

The risk in the project execution as planned is carefully assessed and managed. We base our plans on long standing experience, and we bring together the world's experts in the relevant tools and techniques.

A key feature of this project is the involvement of a wide set of partners from multiple domains. While this ensures complementary coverage of a wide set of skills and provides robustness in different ways, we will have to ensure that all partners work as closely knit team.

Our open source approach means that all our code and outputs are open and visible to anybody at sites like github and bitbucket throughout the project. In particular, it is common for users of computational software to use the most leading edge versions, thus beta-testing code in-between major releases. This results in risk reduction: where our design decision or technical approaches are controversial, this will be detected early by those users, giving the consortium useful feedback to consider.

The project coordinator will, with support from the Coordination Team and Quality Review Board, create a Risk Management Plan D1.3 as part of the Management Work Package, which will be reviewed annually.

Risk	Level with and without mitigation	Mitigation measures
Recruitment of highly qualified staff	High/Medium	Great care was taken identifying pool of candidates to hire from, and coordinating with currently running projects to rehire personnel with strong track record. Typically, we will rehire European postdocs that are currently funded by the Sloan grant to work on Jupyter in California and wish to come back to Europe.
Drowning postdocs under technical work	High/Low	Great care will be taken in distinguishing PhD and postdocs that wish to pursue an academic carrier and full time developers, and assigning to the former tasks with a strong research aspect that will lead to publications (typically in computer science).
Different groups not forming effective team	Medium/Low	Long track record of working collaboratively on code across multiple sites; Aggressive planning of project meetings, work-shops and one-to-one partner visits to facilitate most effective teamwork, combining face-to-face time at one site with remote collaboration.
slmhnlnhfnhs	hsfhs	ghshsh

3.4 Consortium as a Whole

The consortium brings together:

1. Lead or core developers of a cross-section of the major open source computational components for pure mathematics and applications: `GAP` (St. Andrews, Oxford), `LinBox` (Grenoble), `MPIR` (Kaiserslautern), `PARI` (CNRS Aquitaine, Versailles), `Sage` (Orsay, Versailles, CNRS Aquitaine, Oxford, Zürich), `Singular` (Kaiserslautern).

2. Lead developers of a major online mathematical database: `LMFDB` (Warwick, Zürich).

3. Experts in mathematical knowledge management (Bremen).

4. The lead developers of the closest thing currently existing to a Virtual Research Environment for mathematics: `SageMathCloud` (Seattle). Because of the key role of `Sage` in several aspects of the project, and the relevant of `SageMathCloud` as a forerunner of the types of systems we want to build, the participation of this group is essential, even though they are not eligible for Horizon 2020 funding. They have agreed to provide the benefit of their experience on an unfunded basis, since they wish to remain closely involved in all developments in mathematical VREs.

5. Experts and major “promoters” of the `Jupyter` collaborative user interfaces for interactive and exploratory computing in a variety of scientific domains (Southhampton, Simula, Sheffield, Silesia).

6. Lead developers of the Pythran system for automatic conversion of Python to C++ and experts in numerical code optimization/parallelization (Logilab, Grenoble)

7. A company specialized in open-source based Database and Scientific Computing for industry (Logilab); it develops in particular its own virtual environment `Simulagora`.

8. Leading researchers in the sociology of mathematical research and collaboration. In particular the coordinating partner of the “Social Machine of Mathematics” project which has been studying how mathematicians collaborate.

There are many existing points of contact between these groups and communities, although many of them are also new to one another. This, together with the fact that each community is internally collaborative and part of the broader free software community gives us confidence in their ability to work together.

The exact role of each partner in each work package is defined elsewhere, but in general terms:

- Groups 1, 2, and 3 will collaborate be to design the OpenDreamKit VRE architecture – the set of interfaces and standards that allow components to be assembled into bespoke VREs for particular projects or areas. This architecture will be informed by the experience of 7 and 4; by the sociological understanding of 8 and from a **diverse range of real world use cases from all areas of scientific computing, in academia and industry** drawn from their own user bases and contributed by 7 and 4.

  They will be supported in this work by 4, 5 and 6 which bring respectively expertise in the key technologies `SageMathCloud` and `Jupyter` and the cython technology (Seattle).

- All the participants will make use of the OpenDreamKit VRE, providing feedback to the developers and contributing later to the development of demonstrator projects (Objective 8). They will also all participate actively in dissemination (Objective 9) activities.

- Group 5 will host and mentor core `Jupyter` developers to improve this key technology (Objective 2), while 1, 2, and 3 will update their mathematical software components (Objective 4) to comply with the newly developed interfaces. 6 will be a key asset for this work, providing expertise in massive parallelism and HPC, bringing in and further developing the specific `Pythran` optimization technology and providing expertise for development of related technologies in other components.

- Group 7 will further bring in expertise in semantic databases, distribution of large software, and open source based business models.

- Groups 1, 4, and 5 already have strong experience in community building and engagement (Objective 3) for instance through the very active user and developer communities around `GAP` and `Sage`. These communities and the dissemination to them of the availability new free software constitute the primary exploitation route for this project.

3.5 Resources to be Committed

3.5.1 Travel, dissemination and outreach

This community building nature of this grant proposal requires a large number of staff exchanges, work shops with project partners and workshops engaging the wider community in addition to the usual management and project review meetings. For dissemination, we need to target the computer science and focussed community and their conferences as well as the domains benefitting from OpenDreamKit, such as Mathematics and Science.

3.5.2 Travel, dissemination

We use the following guidelines for expected travel expenses: €2200 for attendance of a typical one week international conference outside Europe (including travel, subsistence, accommodation and registration), €1200 for a corresponding conference in Europe, €750 for a one-week visit to project partners, for example for coding sprints and one-to-one visits for particular research. We expect a similar cost per week when hosting visitors. For the 6-monthly project meetings, we expect on average a cost of €400 for travel, accommodation and subsistence.

For a partner site with one investigator and one full time researcher, we expect that that both will attend all of the 9 project meetings that take place every 6 months (cost of 9 * 2 * 400 = €7200), and that the site spends €2000 per year to host external visitors contributing to the project (total €8000). We expect the investigator and research together to do 4 one-week visits to other sites (each at €750, totals €12000).

For dissemination, we expect the researcher to attend on average 1 international conference and 2 European meetings per year (totals €18400) and the investigator to attend one international and one european gathering (totals €13600).

Where there are multiple investigators per site, they will share the travel and associated costs outlined above. Where there are multiple researchers, or researchers not employed for the full 48 months, the travel budget is reduced accordingly.

3.5.3 Outreach costs

We also request €1000 per year per partner to pay for open access publication charges.

We request funds for outreach activities such as workshops that facilitate community building, disseminate best practice and encourages sustained contributions of the community to the project and beyond the lifetime of the funding. For a one-week workshops reaching out to the community, we cost these at about 400 EUR per participant to provide accommodation and catering. A workshop for 15 people will thus cost about €6000. Participants donate their time will fund their own travel. The particular budgeted cost will depend on the local availability of accommodation and will thus vary from workshop to workshop. We use creative means to increase value and improve community building where possible, for example by cooking food ourselves as done in for this recent workshop http://wiki.sagemath.org/days57.

Details are given in the tables below and in the work packages.

.

Summary of staff effort

3.6 Resource summaries for consortium member sites

3.6.1 Resources Université Paris Sud

Please complete section and table. See “Guidelines for completion of partner specific resource summary” in latex comments in H2020/resources.tex.

	Cost (€)	Justification
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 3: Overview: Non-staff resources to be committed at UPS (all in €)

3.6.2 Resources Logilab

Please complete section and table. See “Guidelines for completion of partner specific resource summary” in latex comments in H2020/resources.tex.

	Cost (€)	Justification
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 4: Overview: Non-staff resources to be committed at Logilab (all in €)

3.6.3 Resources Université de Versailles Saint-Quentin

Please complete section and table. See "Guidelines for completion of partner specific resource summary" in latex comments in H2020/resources.tex.

	Cost (€)	Justification
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 5: Overview: Non-staff resources to be committed at UVSQ (all in €)

3.6.4 Resources Université Joseph Fourier

Please complete section and table. See "Guidelines for completion of partner specific resource summary" in latex comments in H2020/resources.tex. UJF requests 12 person months for an engineer (Pierrick Brunet) starting in month 0 and working on task WP5.**T5.7**, 24 person months for another engineer starting in month 12 and working on task WP5.**T5.3**, 12 person months for the lead PI (Clément Pernet) and 6 person months for PI (Jean-Guillaume Dumas). The lead PI will take on all management responsibilities. The engineers will not be employed for the whole project duration, and the PIs will carry out all tasks for the project in the remaining period.

	Cost (€)	Justification
Travel	60,850	Travel (see 3.5.2)
Publication charges	4,000	Open access publication charges (see 3.5.3)
Equipment	24,000	
Other goods and services	30,000	2 Workshops: HPC and Pythran
Audit	?,???	Audit cost
Total	?,???	

Table 6: Overview: Non-staff resources to be committed at UJF (all in €)

3.6.5 Resources CNRS Aquitaine

	Cost (€)	Justification
Travel	107,250	Travel (see 3.5.2)
Publication charges	1,000	Open access publication charges (see 3.5.3)
Equipment	4,000	
Other goods and services	162,500	HPC workshop WP2.**T2.3**, Pari days WP2.**T2.3**, developing countries workshop WP2.**T2.7**
Audit	4,000	Audit cost
Total	278,750	

Table 7: Overview: Non-staff resources to be committed at UB (all in €)

3.6.6 Resources University of Oxford

Please complete section and table. See “Guidelines for completion of partner specific resource summary” in latex comments in H2020/resources.tex.

	Cost (€)	Justification
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 8: Overview: Non-staff resources to be committed at UO (all in €)

3.6.7 Resources University of Sheffield

Please complete section and table. See “Guidelines for completion of partner specific resource summary” in latex comments in H2020/resources.tex.

	Cost (€)	Justification
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 9: Overview: Non-staff resources to be committed at USHEF (all in €)

3.6.8 Resources Southampton

Southampton requests 32 person months for a researcher (expected to start in month 4 of the project), 6 person months for the lead PI (Hans Fangohr) and 2 person months for the co investigator (Ian Hawke). The lead PI will take on all management responsibilities. The researcher will not be employed for the whole project duration, and the PIs will carry out all tasks for the project in the remaining period.

	Cost (€)	Justification
Travel	51,500	Travel (see 3.5.2)
Publication charges	4,000	Open access publication charges (see 3.5.3)
Equipment	10,000	HPC Workstation (6k) to host OOMMF-NB web server, WP4.**T4.10**, and two high performance laptops (2x2k)
Other goods and services	19,600	4 Dissemination workshops overseas (travel for teachers & room hire), WP2.**T2.6**
Audit	5,200	Audit cost
Total	86,300	

Table 10: Overview: Non-staff resources to be committed at Southampton (all in €)

3.6.9 Resources University of St Andrews

St Andrews requests 9.6 person months for the lead PI (Steve Linton), 24 person months for the co-investigator (Alexander Konovalov) and 48 person months for the researcher (Markus Pfeiffer). The lead PI will take on all management responsibilities.

	Cost (€)	Justification
Travel	76,800	Travel (see 3.5.2)
Publication charges	4,000	Open access publication charges (see 3.5.3)
Equipment	15,000	Compute servers for parallel software development and testing (tasks WP5.**T5.2**, WP3.**T3.5**)
Other goods and services	17,200	2 dissemination workshops (room hire and subsistence for external participants; task WP2.**T2.3**)
Audit	4,300	Audit cost
Total	117,300	

Table 11: Overview: Non-staff resources to be committed at USTAN (all in €)

3.6.10 Resources University of Warwick

Warwick requests 24 person months for a researcher expected to start around month 6 of the project to work on WP6, and 3 person months for the lead PI (John Cremona) for WP2 (Management) and WP6. The researcher will not be employed for the whole project duration, and the PI will carry out any remaining tasks for the project. The PI, who is also PI on the LMFDB grant, will be able to use alternative funding for conference attendance, and only requires travel support for project meetings and visiting other sites. The workshop to be hosted will be joint with the LMFDB project and part-funded by the LMFDB grant. Open access publication charges will be met by the host institution.

	Cost (€)	Justification
Travel	9,600	Project meetings and partner site visits; investigator co-funded by LMFDB grant (see 3.5.2)
Equipment	4,000	laptops for investigator and researcher
Other goods and services	12,000	Hosting one workshop co-funded by LMFDB project
Total	25,600	

Table 12: Overview: Non-staff resources to be committed at UW (all in €)

3.6.11 Resources Jacobs University Bremen

Please complete section and table. See “Guidelines for completion of partner specific resource summary” in latex comments in H2020/resources.tex.

	Cost (€)	Justification
Travel	53,600	Travel (see 3.5.2)
Publication charges	4,000	Open access publication charges (see 3.5.3)
Equipment	14,000	for two web/compute servers for WP4.**T4.7**; they need 256 GB RAM each for the math search engine from WP4.**T4.12**
Audit	4,300	Audit cost
Total	75,900	

Table 13: Overview: Non-staff resources to be committed at JacU (all in €)

3.6.12 Resources University of Kaiserslautern

Please complete section and table. See "Guidelines for completion of partner specific resource summary" in latex comments in H2020/resources.tex.

	Cost (€)	Justification
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 14: Overview: Non-staff resources to be committed at UK (all in €)

3.6.13 Resources University of Silesia

Please complete section and table. See "Guidelines for completion of partner specific resource summary" in latex comments in H2020/resources.tex.

	Cost (€)	Justification
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 15: Overview: Non-staff resources to be committed at US (all in €)

3.6.14 Resources University of Zürich

Please complete section and table. See "Guidelines for completion of partner specific resource summary" in latex comments in H2020/resources.tex.

	Cost (€)	Justification
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 16: Overview: Non-staff resources to be committed at UZ (all in €)

3.6.15 Simula Research Labortaory

Please complete section and table. See “Guidelines for completion of partner specific resource summary” in latex comments in H2020/resources.tex.

	Cost (€)	**Justification**
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 17: Overview: Non-staff resources to be committed at Simula (all in €)

3.6.16 University of Washington at Seattle

Please complete section and table. See “Guidelines for completion of partner specific resource summary” in latex comments in H2020/resources.tex.

	Cost (€)	**Justification**
Travel	?,???	Travel (see 3.5.2)
Publication charges	?,???	Open access publication charges (see 3.5.3)
Equipment	?,???	
Other goods and services	?,???	Workshops
Audit	?,???	Audit cost
Total	?,???	

Table 18: Overview: Non-staff resources to be committed at UWS (all in €)

3.7 Resources Summary

Table 19: Overview: Resources to be committed (all in €)

Other direct cost items

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

Management Level Description of Resources and Budget

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

4 Members of the Consortium

4.1 Participants

4.1.1 UPS: UNIVERSITÉ PARIS SUD (FR)

The Université 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 Université 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-Sud 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 (VALS) working on proof assistants (Coq) and another on Human Centered Computing, which will facilitate reaching toward those communities.

The Université Paris-Sud will lead the project (WP1), host the majority of the man power for WP3 Component Architecture, and lead tasks on or around the Sage computational system (WP3.**T3.6**, WP3.**T3.2**, WP4.**T4.1**, WP4.**T4.4**, WP4.**T4.5**, WP5.**T5.6**, ??.??. Finally, it will lead WP2 and in particular host or coorganize many of the community building, training, and dissemination actions.

Curriculum vitae of the investigators

Viviane Pons (leadPI, female, 6 PM) 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.

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.

Florent Hivert (PI, male, 6 PM) Professor at the Laboratoire de Recherche en Informatique, Florent Hivert is a senior researcher in Algebraic Combinatorics with 29 papers in international journals and 15 communications in international conferences.

With 100 tickets (co)authored and as many refereed, Hivert is himself a core Sage developer, with contributions including key components of the Sage infrastructure (documentation, automated test, combinatorics infrastructure, paralellism, . . .), specialized research libraries.

Nicolas M. Thiéry (PI, male, 12 PM) Professor at the Laboratoire de Recherche en Informatique, Nicolas M. Thiéry is a senior researcher in Algebraic Combinatorics with 15 papers published in international journals. Among other things, he is a member of the permanent committee of FPSAC, the main international conference of the domain, and has collaborators in Canada, India, and in the US where he spent three years (Colorado School of Mines, UC Davis). He also coorganized fourteen international workshops, in particular Sage Days, and the semester long program on ”Automorphic Forms, Combinatorial Representation Theory and Multiple Dirichlet Series” hosted in Providence (RI, USA) by the Institute for Computational and Experimental Research in Mathematics.

Algebraic combinatorics is a field at the frontier between mathematics and computer science, with heavy needs for computer exploration. Pioneer in community-developed open source software for research in this field, Thiéry founded in 2000 the Sage-Combinat software project (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[5]. Along the way, he coauthored part of the proposal for NSF Sage-Combinat grant OCI-1147247, and coorganized or taught at a dozen training and dissemination actions (workshops, summer schools, ..), in America, Africa, Europe, India.

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”.

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

Samuel Lelièvre (R, male, 6 PM) Maître de conférences since 2006 at Laboratoire de mathématique d'Orsay, Université Paris-Sud, Samuel Lelièvre is an established researcher in Dynamics and Geometry, with 10 papers published in international journals including Annales scientifiques de l'École normale supérieure, Crelle, GAFA, Geometry and Topology. He participated in three ANR projects, and has collaborators in France, Israel, the UK, the USA. His research in Dynamics and Geometry often involves explicit and experimental approaches, for which he writes code in order to explore combinatorial objects such as square-tiled surfaces, translation surfaces, group actions, group presentations.

He uses and actively promotes Sage since 2010. He is in the top 15 contributors of the Ask-Sage questions and answers site. He coorganized six international meetings including two Sage Days, presented Sage at PyCon-FR-2011 in Rennes, supervised Sage tutorials twice at the GDR-IM yearly school for French PhD students at the interface of Mathematics and Computer Science, and at the CIMPA/ICPAM school Bobo2012 on Discrete Mathematics (Bobo Dioulasso, Burkina Faso, 2012).

Loïc Gouarin (R, male, 5 PM) Research Engineer since 2005 at CNRS and more specifically since 2010 at the Laboratoire de Mathématique d'Orsay, Université Paris-Sud, Loïc Gouarin develops scientific computing software in different fields like Lattice-Boltzmann methods, Stokes solvers for fluid particles interaction, ...

He is also director of the "GdR Calcul" and co-director of the "Réseau Calcul". These two entities form the "Groupe Calcul" of the CNRS whose role is to animate the scientific and high performance computing community in France, in particular by organizing conferences, meetings, and seminars. In this context, he organizes himself 3 to 4 training and development workshops per year, and promotes the use of Python for teaching and research in France.

Technically speaking, and for purely administrative constraints within CNRS, Loïc Gouarin will be attached to the CNRS Aquitaine.

+36

NN (res, 48+36 PM) We will hire two full time experienced software developers to work under the leadership of Nicolas Thiéry on the technical tasks pursued by Paris Sud, in particular in WP3 Component Architecture and WP4 User Interface. When relevant, the mentoring will be complemented by Luca De Feo (UVSQ), or experts of the tasks at hand from the larger community. The fellow will have a strong software engineering experience, ideally in the Python ecosystem. We further require good communication and team working skills, in particular to work in tight collaboration with international open-source developer communities.

NN (res, 24 PM) We will hire an experienced part time project manager to help with the overall management of OpenDreamKit.

Publications, achievements

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. 500 tickets contributed to `Sage`.

Previous projects or activities

1. Home of six Sage Days (week-long training and development workshops).
2. Coorganizer of six other Sage Days.
3. Founder and regular organizer of a bimonthly `Sage` User Group meeting in the greater Paris area.

Significant infrastructure

The Université Paris Sud hosts the lead developers of the open source cloud infrastructure `Stratuslab` and its reference infrastructure (400 cores). The participants are regular users of this infrastructure, and in close contact with the developers.

Paris Sud also hosts the WILDER platform, an experimental wall-sized high-resolution interactive touch-screen for conducting research on collaborative human-computer interaction and the visualization of large datasets.

4.1.2 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.

Curriculum vitae

Florent Cayré (PI, 2 PM) Engineer with a Master Degree from École Centrale de Paris (top French engineering school), Florent Cayré spent six years in SNECMA as an engineer conceiving the numeric tools for the turbines design and then was head of the group “Méthodes de conception de turbines ; aérothermique et combustion”: collaborative R&D programs, development and integration of various numeric tools (Python, C++, C, Fortran). He co-funded the SecondWeb company: development of complex Web applications based on CubicWeb platform.

Head of the science computing department of Logilab since 2012,x Florent is responsible for team management, strategic vision, projects monitoring, technical expertise, etc. He developed several tools for defining and managing computations, and producing enhanced result reports through IPython notebooks.

Olivier Cayrol (lead PI, 4 PM) Engineer with a Master Degree from École Centrale de Lyon (top French engineering school) Olivier Cayrol spent three years at the R&D department of PSA-Peugeot Citroën, as developer and project manager on the modelling and simulation of electronic embedded car control devices.

Co-founder and deputy-CEO of Logilab, he designed and developed the system that generates the Logilab's documents from ReST data sources. This system is based on several free softwares such as Sphinx or reportlab, and defines numerous extensions for answering the specifc needs.

Serge Guelton (R, male, 12 PM) Core dev of Pythran ...

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

4.1.3 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 (leadPI, male) 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 (R, male) 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.

4.1.4 UJF: Université Joseph Fourier (FR)

The Université Joseph Fourier (UJF) is a research intensive university in an international and high tech environment with $16\,600$ full time students including $1\,200$ in a doctoral program, $1\,500$ lecturers and researchers and $1\,500$ administrative and technical staff. Ranked 150th at the latest Shangaï ranking, 5th french university, and top 75 natural sciences and mathematics. UJF laboratories have successfully contributed to the FP7 european with 105 projects, 26 of them being coordinated by UJF. Under H2020, UJF continues to actively participate with already 4 projects.

The UJF partner gathers two teams from the Laboratoire Jean Kuntzmann (CASYS team, with Jean-Guillaume Dumas) and the Laboratoire d'Informatique de Grenoble (MOAIS, Clément Pernet). 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 the exact arithmetic library `Givaro`, the exact linear algebra libraries `fflas` and `LinBox`, and the parallel programming runtime XKAAPI. Relying on this expertise, the UJF site will be coordinating the tasks related to High Performance Mathematical Computing, forming Work Package 5: WP5.**T5.1**, WP5.**T5.2**, WP5.**T5.3**, WP5.**T5.4**, WP5.**T5.5**, WP5.**T5.6**, WP5.**T5.7**, WP5.**T5.8**. It will also be in charge of the task WP5.**T5.3** on the development of high performance exact linear algebra with the `LinBox` library.

Curriculum vitae

Clément Pernet (leadPI, male, 12 PM) 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.

Jean-Guillaume Dumas (PI, male, 6 PM) 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.

Pierrick Brunet (R, male, 12 PM) 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.

Unknown (res, 24 PM) Full time developer on `LinBox`, and its integration into `Sage`. The recruited person should master C++ programming and library development in general. A good knowledge on scientific computing and in particular linear algebra is also required.

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. `Pythran` takes advantage of mutli-cores and SIMD instruction units and, thanks to type inference, it requires little annotations. Its runtime supports a growing subset of the `Numpy` package.

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.

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.

4.1.5 CNRS: CNRS Aquitaine (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

Vincent Delecroix (leadPI, male, 12 PM) CNRS researcher at the LaBRI (Bordeaux, France) since october 2013, Vincent Delecroix is a junior researcher in Dynamical Systems with strong links with Combinatorics and Number Theory. He published 7 articles in international journals with several collaborators around the world (England, Mexico, United-States).

Since 2010 he is an important contributor of Sage with 30 tikets authored and around 50 reviewed. He organized several Sage days and Sage workshops in Marseille, Orsay, Bobo Dioulasso (Burkina Faso), Bordeaux.

Karim Belabas (PI, male, 12 PM) 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).

Bill Allombert (PI, male, 6 PM) Research ingenier in Bordeaux. He has a great expertise in algorithmic number theory and is one of the main developer of PARI/GP.

He also contributes to the software GAP and the debian project.

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.
5. V. Delecroix *Cardinality of Rauzy classes*

ix ”Cardinality of Rauzy classes” Ann. Inst. Fourier, 63 no 5 (2013), p. 1651-1715.

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

Two center for computations: Plafrim and Avakas.

4.1.6 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

Ursula Martin (PI, female, 2 PM) 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.

Edith Elkind (PI, female, 2 PM) 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.

Dmitrii Pasechnik (leadPI, male, 24 PM) 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

Oxford has world-class computational facilities, including numerous HPC clusters and a dedicated centre, Advanced Research Computing, to support HPC users. Another dedicated facility, Oxford's e-Research Centre, facilitates digital research and drives innovation in technology. Last but not the least, the library of Oxford University is one of the most complete libraries in the UK.

4.1.7 USHEF: University of Sheffield (UK)

The University of Sheffield is a leading Reasearch University in the United Kingdom that was ranked 69th in the World in the most recent 2014 QS World University Rankings and was ranked in the top ten for the most recent UK wide research asssment exercise. Professor Lawrence is based in The Sheffield Institute for Translational Neuroscience (SITraN) and the Department of Computer Science at the University of Sheffield have a unique partnership based on two shared professorial appointments, Professor Winston Hide and Professor Neil Lawrence. SITraN is a world leading research centre for neurodegenerative disease, located in a purpose built building on the University of Sheffield campus adjacent to the Medical School at Sheffield Teaching Hospitals NHS Trust Royal Hallamshire Hospital.

The Department of Computer Science was ranked 5th across UK departments in “Research Quality” by the recent UK-wide Research Evaluation Framework. It has a particular history of working with data with internationally leading groups in Machine Learning, Speech and Language Processing.

Curriculum vitae

Neil Lawrence (leadPI, male, 6 PM) Neil Lawrence received his bachelor's degree in Mechanical Engineering from the University of Southampton in 1994. Following a period as an field engineer on oil rigs in the North Sea he returned to academia to complete his PhD in 2000 at the Computer Lab in Cambridge University. He spent a year at Microsoft Research in Cambridge before leaving to take up a Lectureship at the University of Sheffield, where he was subsequently appointed Senior Lecturer in 2005. In January 2007 he took up a post as a Senior Research Fellow at the School of Computer Science in the University of Manchester where he worked in the Machine Learning and Optimisation research group. In August 2010 he returned to Sheffield to take up a collaborative Chair in Neuroscience and Computer Science.

Neil's main research interest is machine learning through probabilistic models. He focuses on both the algorithmic side of these models and their application. He has a particular focus on applications in personalized health and computational biology, but happily dabbles in other areas such as speech, vision and graphics.

Neil was Associate Editor in Chief for IEEE Transactions on Pattern Analysis and Machine Intelligence (from 2011-2013) and is an Action Editor for the Journal of Machine Learning Research. He was the founding editor of the JMLR Workshop and Conference Proceedings (2006) and is currently series editor. He is Programme Chair for AISTATS 2012 and has served on the programme committee of several international conferences. He was an area chair for the NIPS conference in 2005, 2006, 2012 and 2013, Workshops Chair in 2010 and Tutorials Chair in 2013. He was general chair of AISTATS in 2010 and AISTATS Programme Chair in 2012. He is Program Chair of NIPS in 2014.

Neil is a strong advocate of open source software in machine learning and has given many invited talks on the subject. Since 2004 his research group has made all their implementaitons available, most recently using Python and IPython as the main medium for communicating their work. Their Gaussian process python software framework[6] is becoming a standard platform for research in these methods and underpins a series of Summer Schools and four day road shows that Neil has led in the area.[7]

Michael Croucher (R, male, 6 PM) is a Research Software Engineer at The University of Sheffield. He received his bachelor's degree in Theoretical Physics from The University of Sheffield in 1999 and completed his Theoretical Physics PhD there in 2005. He subsequently took up a research-support post in The University of Manchester's IT Services department before being appointed as Head of Application Software Support for the Faculty of Engineering and Physical sciences.

Michael is passionate about improving the quality of research software. He enables researchers to ask larger and more complex research questions by improving the software they develop. By teaching and demonstrating fundamental software engineering principles, he assists academic colleagues in producing robust, reproducible, fast and correct software.

He achieves these aims via a number of means:

Consultancy: He works directly on research code written in various languages. For a sample of recent client testimonials, see http://www.walkingrandomly.com/?page_id=5122

Outreach: He is the author of http://www.walkingrandomly.com/ - a blog focused on mathematics and scientific computing with over 400,000 unique visitors a year. The associated twitter account, @walkingrandomly, has almost 3000 followers. He is a fellow of the Software Sustainability Institute, an organisation that promotes and supports research software engineering.

Education: He has taught programming and Software Carpentry classes to hundreds of researchers and uses his contacts with education and industry to arrange specialised teaching events relating to research software.

Mentoring: He acts as a 'code-coach' to new researchers, providing code reviews and private tutorials.

[6] https://github.com/SheffieldML/GPy
[7] http://ml.dcs.shef.ac.uk/gpss/

Vendor liaison: He has strong relationships with several vendors of scientific software including Mathworks, Wolfram Research, Maplesoft and NAG. These relationships have led to many fruitful collaborations between them and academic colleagues.

High Performance Computing: He has been involved with teams that develop and support Institutional HPC systems such as the Manchester University Condor Pool - a 3000+ CPU core facility made by utilising spare time on hundreds of desktop PCs. In this team, his primary role was to assist researchers in transitioning their workflow from the desktop to the Condor system.

Publications, products, achievements

1. N. Fusi, C. Lippert, N. D. Lawrence and O. Stegle. (2014) ”Warped linear mixed models for the genetic analysis of transformed phenotypes” in Nature Communications 5 (4890)
2. J. Hensman, M. Rattray and N. D. Lawrence. (2014) ”Fast nonparametric clustering of structured time-series” in IEEE Transactions on Pattern Analysis and Machine Intelligence
3. M. A. Álvarez, D. Luengo and N. D. Lawrence. (2013) ”Linear latent force models using Gaussian processes” in IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (11), pp 2693–2705
4. N. Fusi, O. Stegle and N. D. Lawrence. (2012) ”Joint modelling of confounding factors and prominent genetic regulators provides increased accuracy in genetical genomics studies” in PLoS Computat Biol 8, pp e1002330

Previous projects or activities

1. Organisers of the Gaussian Process Summer Schools (three 3-day workshops on Gaussian process models in python and the IPython notebook).
2. Organisers of five Gaussian Process and Data Science Road Shows (educating on data science and Gaussian process models in Uganda, Colombia, Italy, Australia and Kenya) Each workshop is 3-4 days long.

Significant infrastructure

1. The Sheffield Institute for Translational Neuroscience is a 18 million pound world leading institute for research into neurodegenerative disorders. It houses clinicians, biologists and computationalists under a single roof and contains the Sheffield Microarray and Next Generation sequencing Core Facility. The institute provides an exemplar of how mathematical ideas can be rapidly translated to analysis through provision of appropriate software frameworks.

4.1.8 USO: University of Southampton (UK)

The University of Southampton (USO) 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. The University of Southampton 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. 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. The university has a strong track record of working in European projects, especially within the Framework Programme. The EC 6th FP7 Monitoring Report ranked USO 17th out of all higher and secondary education organisations for number of FP7 participations during 2007-2012. Throughout the FP7 USO has received € 132M in research grants and has been involved in 319 projects, including 63 ICT and 8 INFRASTRUCTURES Collaborative Projects.

The Faculty of Engineering and the Environment (FEE) consists of 370 research postgraduate students and 340 academic and research staff, and 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. The faculty 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 the context of this proposal, Southampton has long standing experience in high performance computer simulation to advance science and engineering, with significant hardware infrastructure, a critical mass of several hundred researchers working in the field, and a dedicated doctoral training center in computational modelling. Southampton's main tasks in this project are the extension and application of the `Jupyter` Notebook technology in computational materials research to provide a virtual research environment demonstrator for a large research community.

Curriculum vitae

Hans Fangohr (leadPI, male, 6 PM) 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 prodives 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 ($\pounds 1000$) 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.

Ian Hawke (PI, male, 2 PM) Ian Hawke is a lecturer in Applied Mathematics at the University of Southampton and a co-director of the € 12m EPSRC Centre for Doctoral Training in Next Generation Computational Modelling. An expert in nonlinear simulations of relativistic matter and numerical techniques, he has taught numerical methods in many contexts for ten years. He has worked on IPython (`Jupyter`) Notebooks in education, particularly as an instructor on the "Practical Numerical Methods in Python" MOOC, which builds on other open technologies including OpenEdX and GitHub. The initial author of the "Whisky" relativistic hydrodynamics code, he has been a contributor to and maintainer of a range of projects used across the numerical relativity community, including the Einstein Toolkit, the Cactus infrastructure and the Carpet mesh refinement code. His recent research has concentrated on numerical methods for relativistic matter beyond ideal fluids, including modelling sharp transitions and surfaces, relativistic elasticity, and the first nonlinear simulations of relativistic multifluids.

NN (R, 32 PM) 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.

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)

4.1.9 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. Now GAP and its packages comprise over 1 million lines of C and 1.25 million lines of GAP code. The system is distributed freely under the GPL2, and our records show that it has been installed in at least 3000 sites and cited in several thousands publications (see http://bit.ly/gap_citations and http://www.gap-system.org/Doc/Bib/bib.html). GAP was used in landmark computations such as the "Millennium Project" to classify all groups of order up to 2000 and the classification of the over 10^{19} semigroups of order 10. It is designed to be natural to use for mathematicians; to be powerful and flexible for experts and to be freely extensible so that it can encompass new mathematics. These objectives have been met and GAP was awarded the ACM/SIGSAM Richard D. Jenks Memorial Prize for Excellence in Software Engineering applied to Computer Algebra in 2008.

Nowadays, as one of the centres of GAP development, CIRCA has excellent contacts with developers and users worldwide. Particularly relevant to this proposal are the Singular and homalg groups at Kaiserslautern, Soicher at QMUL, Praeger and others at UWA, Cooperman and his students at NEU and the other GAP centres in Aachen, Fort Collins and Braunschweig.

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 (leadPI, male, 10 PM) 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[8]. 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 (PI, male, 24 PM) 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.

Markus Pfeiffer (R, male, 48 PM) is a Research Fellow at the University of St Andrews active in the School of Computer Science and the School of Mathematics and Statistics. His unusual breadth of knowledge encompasses expertise in formal language theory, decidability, and algebra, as well as practical computation and programming languages. Since receiving his PhD in 2013 he has been an active researcher in semigroup theory as well as contributing to the GAP system both as a package author, and as a core system developer.

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.
2. V. Janjic, C.M. Brown, M. Neunhöffer, K. Hammond, S. Linton, H-W. Loidl. Space exploration using parallel orbits: a study in parallel symbolic computing. in Parallel Computing: Accelerating Computational Science and Engineering (CSE). vol. 25, Advances in Parallel Computing, IOS Press, 2013, p. 225–232.

[8] Applicable Algebra and Error Correcting Codes

3. A. Konovalov and S. Linton. SCSCP — Symbolic Computation Software Composability Protocol, Version 2.1.4; 2013. GAP package (http://alexk.host.cs.st-andrews.ac.uk/scscp/).
4. V. Slavici, D. Kunkle, G. Cooperman, S. Linton. An efficient programming model for memory-intensive recursive algorithms using parallel disks. In ISSAC 2012: Proceedings of the 37th International Symposium on Symbolic and Algebraic Computation. New York, ACM Press, 2012, p. 327–334.
5. R. Behrends, A. Konovalov, S. Linton, F. Lübeck, M. Neunhöffer. Towards high-performance computational algebra with GAP. Proceedings of the Third International Congress on Mathematical Software: Kobe, Japan, September 13-17, 2010. LNCS 6327, Springer, 2010, p. 58–61.

Previous projects or activities

1. **Multidisciplinary Critical Mass in Computational Algebra and Applications** EP/C523229 2005–2010, £1.1m. Through a range of subprojects this grant developed CIRCA as a broad centre of excellence in computational algebra. We extended GAP, developed new algorithms, and opened up new applications in AI, combinatorics and physics. The project produced over 200 refereed publications in a huge range of venues, and delivered, as intended, a sustainable step change in the scale and breadth of CIRCA's research.
2. **SCIEnce: Symbolic Computation Infrastructure for Europe** (FP6 eRII3-CT-026133) 2006–2011, € 3.2m, coordinator. This project tackled the fragmentation of the European community of researchers in, and users of, symbolic computation. Among the nine partners were four major system developers (of GAP, Maple, MuPAD and KANT), an international research institute (RISC-Linz) and other groups with specialist expertise. Project activities included symbolic web services, system composability, symbolic grid and cloud computing and a program of visits, workshops and summer schools. One important outcome was a new protocol SCSCP, now used well beyond the original project.
3. **HPCGAP: High Performance Computational Algebra and Discrete Mathematics** EP/G055181 2009–2014, £1.5m, four sites. This project aimed at reengineering GAP to support simple, safe and efficient parallel programming on a range of systems from multicore laptops and desktops, through departmental and university clusters to HPC systems. By the end of the project, we had adapted a complex system including a language runtime and over 400 000 lines of interpreted code to enable safe and efficient parallel programs. The proposed project is very timely as the multi-threaded version of GAP is becoming mainstream, and users and package developers need training and support to parallelise their code.

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.

4.1.10 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

John E. Cremona (leadPI, male, 3 PM) 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 work package "Effective Cohomology Computations" in the second, responsible for several deliverables. 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 including the LMS Journal of Computation and Mathematics and the Journal of the Foundations of Computational Mathematics (FoCM). He has supervised 16 PhD students, a dozen Masters students, two EU-funded Marie-Curie fellows and currently has three EPSRC-funded postdoctoral research assistants working for the LMFDB project. Cremona has given over 30 invited conference addresses and seminars in 9 countries in the last 10 years; most recently he was a Plenary Speaker at the 2014 FoCM meeting in Montevideo, where he spoke about the LMFDB project to a wide international audience of computational mathematicians.

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-used 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.

NN (R, 24 PM) We will hire a computational support technician to work at Warwick, under the leadership of Professor John Cremona, on the tasks of Work Package 6 (Data/Knowledge/Software-bases). He or she will develop closer integration between Sage and the LMFDB, as a prototype for similar integration between mathematical software and databases. He or she will have a background in computational science, with experience of software engineering, including front-end web development (HTML5, CSS), will be proficient in Python and web-development libraries (flask, jinja2), and have knowledge of SQL and NoSQL databases (SQLite, MongoDB). We further require the person to have good communication and team-working skills, to be able to communicate technical details to casual programmers and able to prioritise and delegate tasks. Experience with Sage and other mathematical software will be an advantage.

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, organised by John Cremona with Samir Siksek, included six research workshops and a summer school “Number Theory for Cryptography” and raised the international profile of the number theory group substantially.
3. As well as workshops for the LMFDB project, John Cremona has co-organised a Flint-Sage-Days workshop at Warwick with William Hart (now at Kaiserslautern).

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.

4.1.11 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 [12]) 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

Michael Kohlhase (leadPI, male, 6 PM) 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).

Florian Rabe (PI, male, 6 PM) 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.

Christian Maeder (Res, male, 24 PM) 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.

Mihnea Iancu (JRes, male, 24 PM) 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 [13] and redeveloped its formal core in the OMDoc/MMT format [24]. The latter has been implemented in the MMT system [23, 24] 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[15, 14].

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 [11] as well as automatic annotation using linguistic techniques [9]. 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 [16, 17, 10].

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

4.1.12 UK: UNIVERSITY OF KAISERSLAUTERN (DE)

4.1.13 University of Kaiserslautern

Prof. Dr. Wolfram Decker (PI, male) 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 (R, male) 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.

4.1.14 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

Marcin Kostur (leadPI, male, 12 PM) 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.

Jerzy Łuczka (PI, male, 12 PM) Prof. Dr. Jerzy Łuczka (his web page: http://zft.us.edu.pl/luczka) is a full professor of physics at the University of Silesia (Katowice, Poland). He heads the Department of Theoretical Physics.

He published more than 150 papers in journals (all on ISI Master Journal List) which have been cited almost 2000 times.

He is Editor of European Physical Journal B, Chairman of the Statistical and Nonlinear Physics Division (European Physical Society), Fellow of the Institute of Physics (United Kingdom) and Outstanding Referee (American Physical Society). He was Co-director of the NATO Advanced Research Workshop "Stochastic Systems. From randomness to complexity", 2002, Erice (Italy) and Member of the Steering Committee of the program : "Stochastic Dynamics: Fundamentals and applications" (European Science Foundation), 2003-2008. He received the DAAD research fellowship (Forschungsaufenthalte für Hochschullehrer und Wissenschaftler) 1995, 2009 and 20012. He was a leader of several Polish and two German-Polish grants. He has collaborators in Germany, Italy and Spain. He has also co-organized international conferences.

Łuczka's research interests lie in areas of stochastic processes in physics, quantum open systems, transport phenomena, physical fundamentals of quantum information. He has teaching experience with Sage in physics, biophysics and econophysics.

Jan Aksamit (PI, male, 12 PM) got his PhD in 1982 and worked at University of Silesia as research assistant, lecturer and senior lecturer. His skills combine forty 40-years of experience in teaching algebra, classical and quantum mechanics, quantum information, statistical physics, and mathematical methods of physics with proficiency in computing. He has actively participated in the project iCSE (innovative Computing in Science Education), where he created Sage enhanced textbook of Linear Algebra (polish only: http://visual.icse.us.edu.pl/LA) using modern interactive technologies. In this project his main task will be to pursue to exploit the vast lecturing experience for creation of interactive demonstrators of VRE.

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.

4.1.15 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

Paul-Olivier Dehaye (leadPI, male, 6 PM) 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.

4.1.16 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.

Curriculum vitae

Hans Petter Langtangen (leadPI, male) 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.

4.1.17 UWS: University of Washington at Seattle (US)

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

Curriculum vitae

William Stein (R, male, 0 PM) 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.
[**science-project**]

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

No third parties involved.

5 Ethics and Security

5.1 Ethics

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 pp. 43, 45).

[2] http://math.nist.gov/oommf/oommf_cites.html (cit. on p. 45).

[3] (Cit. on p. 60).

[4] 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. 60).

[5] *CVE-2014-6271*. (cit. on p. 60).

[6] *Docker – An open platform for distributed applications for developers and sysadmin*. hrefhttp://docker.comhttp://docker.com. 2015 (cit. on p. 48).

[7] 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. 60).

[8] 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. 43).

[9] 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. 93).

[10] 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. 93).

[11] 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. 93).

[12] *Knowledge Adaptation and Reasoning for Content*. URL: http://kwarc.info (visited on 05/12/2011) (cit. on p. 93).

[13] 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. 93).

[14] 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. 93).

[15] 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. 93).

[16] 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. 93).

[17] *MathHub.info: Active Mathematics*. URL: http://mathhub.info (visited on 01/28/2014) (cit. on p. 93).

[18] *MathOverflow*. URL: http://mathoverflow.net (visited on 09/08/2011) (cit. on p. 59).

[19] N. Nisan. “Introduction to mechanism design (for computer scientists)”. In: *Algorithmic game theory*. Cambridge Univ. Press, Cambridge, 2007, pp. 209–241 (cit. on p. 60).

[20] *Nmag — Guided tour*. http://nmag.soton.ac.uk/nmag/0.2/manual/html/guided_tour.html (cit. on p. 48).

[21] *Nmag micromagnetic simulation suite*. http://nmag.soton.ac.uk (cit. on p. 43).

[22] *Notebook based interactive GUI for common simulation requirements*. https://www.youtube.com/watch?v=RhMHgQbP__A. 2014 (cit. on p. 48).

[23] F. Rabe. *The MMT Language and System*. URL: https://svn.kwarc.info/repos/MMT (visited on 10/11/2011) (cit. on p. 93).

[24] 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. 93).

[25] J. Rehmeyer. *Massively Collaborative Mathematics*. Apr. 1, 2010. URL: http://www.siam.org/news/news.php?id=1731 (visited on 01/08/2015) (cit. on p. 59).

[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. 60).

[27] *The polymath blog*. URL: http://polymathprojects.org/ (visited on 07/12/2010) (cit. on p. 59).

[28] *Trac Open Source Project*. (cit. on p. 60).

[29] *VirtualBox*. hrefhttp://virtualbox.orghttp://virtualbox.org. 2015 (cit. on p. 48).