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

Date of Preparation: January 13, 2015

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 (coordinator)	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

Abstract

.

OpenDreamKit will deliver a flexible toolkit that will make it easy to set up 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.

This will be achieved by a Europe-wide steered-by-demand collaboration that assimilates a leading body of mathematicians, computational researchers, and software developers, all with a long track record of delivering innovative open source software solutions meeting the diverse needs of their respective communities.

By concentrating the efforts on improving and unifying existing general purpose building blocks, OpenDreamKit will simultaneously maximize sustainability and impact, with a broad range of 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, security and society overall are driven though innovation. The key tools for innovation are mathematical knowledge and algorithms. Our global positioning system (GPS) needs relativistic mathematics, our mobile phones are allocated frequencies through combinatorial optimization, the combinatorics of our genome yields clues to curing rare diseases, the privacy of our communications depends on cryptographic protocols steeped in number theory, and our national security is reliant on the mathematical analysis of increasingly complex networks. Engineering, Science and Business innovations that enrich society and mankind are made possible through mathematical foundations which are often developed long before their potential applications. Reciprocally, modern mathematical research is increasingly accelerated by and enabled through collaborative tools, computational environments and online databases. These digital tools have the potential to revolutionise the way 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 building of specific *Virtual Research Environments* (VREs) for particular tasks and communities.

We will achieve this by focusing on a *toolkit of software components* from which *tailored VREs can be assembled flexibly* to cater for the diversity and evolution of needs in mathematics, science and engineering. We are at a critical point providing an opportunity to do so: collaborative tools for code sharing (e.g. `github`) now allow us to bring together very large communities of open source code developers.

Simultaneously the last decade has witnessed the emergence of fundamental open source building blocks, at the forefront of which are computational components such as the general purpose mathematical software system `Sage` and the interactive computing environment `Jupyter` (successor of `IPython`). Throughout this project we will reuse and extend open source code, and OpenDreamKit will benefit from future open source contributions during and beyond the lifetime of the project. By unifying tools with overlapping functionality, such as `Jupyter` and `Sage` with their notebooks, we focus the effort of the computational community onto OpenDreamKit, producing additional economies of scale. Finally, thanks to the “by users for users” model, the development will be steered by the actual needs of the community.

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 collaborative workspace. The basic units are executable documents, i.e., data- and code- driven narratives that combine live code, equations, text, interactive dashboards and other rich media. Potential applications include active scientific logbooks, papers, lecture notes, etc., covering the whole lifecycle of a mathematical research project.

This will enable step changes in effective research, research communication, and reproducibility in computational mathematics and science. It will further provide end-to-end toolchains that link fundamental mathematics to domain specific specialised computation, thus bridging the gap between fundamental research and technology, and paving the way towards faster application, exploitation and commercialisation of basic research.

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

The OpenDreamKit team is a Europe-wide collaboration that brings together a leading body of mathematicians and transdisciplinary computational researchers, with an extensive track record of delivering innovative open source software solutions.

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, though the primary target users are *researchers in mathematics*, the set of beneficiaries extends to workers in scientific computing, physics, chemistry, biology, engineering, medicine, earth sciences and geography, social sciences and finance, and includes 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 SMEs.

1.1 Objectives

Our research will cover a wide variety of aspects. For the construction of the OpenDreamKit virtual environment toolkit we consider component architecture, user interfaces, deployment frameworks and standardization of software systems. We also consider the social and technical aspects that will ensure the utility of the virtual environment framework including: publication, open source tools and model development, data archiving and reproducibility of mathematical experiments.

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 set up custom, collaborative, *Virtual Research Environments* tailored to their specific needs, resources and workflows. The VRE should support 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; exploiting and supporting the growing ecosystem of computational tools.

Aim 4: Maximize sustainability and impact in mathematics, neighbouring fields, and scientific computing.

We will achieve our aims through nine objectives, as listed below.

Objective 1: To develop and standardize an architecture allowing combination of mathematical, data and software components with off-the-shelf computing infrastructure to produce specialized VRE for different communities. This primarily addresses aim 2, thereby contributing to aims 1 and 3.

Objective 2: To develop open source core components for VRE where existing software is not suitable. These components will support a variety of platforms, including standard cloud computing and clusters. This primarily addresses Aim 2, thereby contributing to Aim 1 and 3.

Objective 3: To bring together research communities (e.g. users of `Jupyter`, `Sage`, `Singular`, and `GAP`) to symbiotically exploit overlaps in tool creation building efforts, avoid duplication of effort in different disciplines, and share best practice. 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: Ensure that our ecosystem of interoperable open source components is *sustainable* by promoting collaborative software development and outsourcing development to larger communities whenever suitable. This fulfills part of Aim 3 and 4.

Objective 6: Promote collaborative mathematics and science by exploring the social phenomena that underpin these endeavours: how do researchers collaborate in Mathematics and Computational Sciences? What can be the role of VRE? How can collaborators within a VRE be credited and incentivised. This addresses part of Aim 3, 1, and 2.

Objective 7: Identify and extend ontologies and standards to facilitate safe and efficient storage, reuse, interoperation and sharing of rich mathematical data whilst 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: Promote and disseminate OpenDreamKit to the scientific community by active communication, workshop organization, and training in the spirit of open-source software.

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 (e.g. GAP, `PARI/GP`, `Sage` or Singular) and valuable on-line services (e.g. the Encyclopedia of Integer Sequences[1] and the ATLAS of Finite

[1] `http://oeis.org`

Group Representations[2]). In building these systems, mathematicians have gained strong experience in collaborative software development, with pioneering work and continuing leadership in Europe.

A number of approaches to linking these resources have been developed, such as the SCSCP protocol from the Framework 6 SCIEnce project[3], 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.

In Objective 1 we will *design an architecture* which will allow existing mathematical software systems off-the shelf non-mathematical tools and a small number of new components to be flexibly combined to produce versatile VRE that will support collaborative mathematical research throughout its entire life-cycle. This will include software APIs, standards, and 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 VRE fit the ways that mathematicians actually work.

Our research covers all aspects of the ecosystem, both technical (software development models; user interfaces; virtual environments; deployment frameworks; novel collaborative tools; component architecture; design; standardization of software components and databases) and social/collaborative (publication; data archive; reproducibility of experiments; development models; development tools; social aspects).

It will build on the 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 mathematical capabilities of our software will come from existing or updated open source mathematical systems (e.g. the GAP Library for computational group theory and PARI/GP for number theory). And generic services such as storage, version control (e.g. github), authentication and resource accounting will come from off-the-shelf components building on standard infrastructures.

However, core *new tools* will need to be developed or adapted. 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 notation.

Objective 3: Community building across disciplines

Open source development is most efficient when the load is shared as widely as possible. However, across different communities a lack of communication can mean that good ideas are re-invented or re-implemented, when a shared resource would be more efficient. By fostering a more *cross-disciplinary* community, sharing tools where possible and by creating generic tools for wide distribution we will reduce duplication of effort. This will lead to high *quality* software that is more *sustainable*. The maintenance and development effort can be focused on one tool rather than a disparate spread of codebases. This will ensure innovative ideas and best-practice are shared more effectively, increasing research productivity.

While each of the communities such as the developers of Sage, Singular, and GAP need somewhat special features for their research, they are united through being (i) focussed on mathematical challenges, and (ii) needing a computational workflow. IPython and the Jupyter Notebook are used widely in science and engineering. These communities are based on (iii) applications of mathematics that also require computational workflows for collaborative research and dissemination. These three common attributes distill the requirements for core features of the VRE described in this 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

Our vision leverages the community's 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 interoperate 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. The rapid takeoff of Sage in the last decade has proven the viability of the "developed by users for users" model for general purpose systems in pure mathematics. Sage development is driven by an international community of about 150 active developers, many based in universities. Most activities are funded indirectly

[2] http://brauer.maths.qmul.ac.uk/Atlas/v3/
[3] http://www.symbolic-computing.org

by research grants targeted at specific development in mathematics, where the software component is often an indirect outcome.

This somewhat piecemeal approach is enabled by (i) reusing existing components wherever possible (including hundreds of specialized open source math libraries and the `Python` programming language, with its developers' tools and huge library) (ii) spinning off software development (e.g. the `Cython` compiler) to larger communities whenever possible and (iii) carefully designing the development workflow.

However, critical long-term non-mathematical features: e.g. portability; modularization; packaging; user interfaces; large data; parallelism; outreach toward related software, have lagged behind. Principally this is because these components are not credible indirect developments of stand alone projects. They need to be assigned to a small group of full time developers. Regular funding is also needed to improve dissemination of the toolkit to ensure the benefits of more productive pipeline of research are felt by the wider research community that is critically dependent on mathematical developments. This grant will pump prime that process enabling longer term planning and a more structured approach to component development and assimilation.

We envisage that with the growth of the user base a core group of institutions or companies will hire full-time developers to support the 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. At the scale of a large university or company, the cost of software licenses for commercial equivalents to `Sage` can easily outstrip the cost of a small team of developers. Our proposal for VRE goes beyond any commercial software provision and bridges the gap between end users and developers that typically delays the advance of commercial systems.

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. OpenDreamKit will save the mathematics community from duplication of effort, by first outsourcing the development of the user interface of each computational component to `Jupyter`, ensuring that `Jupyter` stands up to the stringent needs of the community. `Jupyter`'s large industrial and academic user base will benefit from these contributions, but is not reliant on the mathematics community (either in development effort or for funding) to remain sustainable.

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

Objective 6: Engineering social interactions in open source VRE

Scientists interact in the process of scientific discovery in a variety of ways. In particular, researchers in mathematics and adjacent theoretic disciplines often refer to minds of collaborators as some sort of laboratories. With the advent of internet the volume of scientific communication increased by orders of magnitude. Recent successful massively collaborative online projects to attack mathematical problems, known as *Polymath*, initiated by Gowers and Tao, were not feasible 20 years ago. However, not all initiatives are successful. There is a social aspect to interactions of this type that is critical to a productive collaboration.

In many ways the process of designing, development and maintenance of an open-source VRE, in which mathematics and algorithms play a key role, closely resembles *Polymath* efforts. Thus OpenDreamKit VRE are perfect objects to study, as there is plenty of data to analyse, and opportunity to tweak the development workflow to obtain more relevant data, if needed.

Social aspects of interactions have become a focus of attention of a burgeoning field of *algorithmic game theory*, which provides tools to engineer environments where all participants are incentivised to contribute to the common good. Finding an optimal way to allocate reputation scores to participants of an online trading platform, such as Ebay, is just one example where these tools are used. We will investigate optimal ways to allocate reputation scores to developers and users of on open source VRE for their contributions, to facilitate the "mutual crowdsourcing" that is taking place as the VRE toolkit evolves, using the OpenDreamKit VRE as a testing ground.

Another set of tools deals with questions of stability of coalitions and collective decision making. These are applicable to the questions of stability of the community of developers of open source VRE. Not all open source projects achieve a stable, sustainable, status: forks are created, developers leave, community interest dwindles. This results in a waste of resources. We develop tools for improving the stability of open source projects using cooperative game theory. These tools will be applied to our own development efforts in OpenDreamKit VRE.

Objective 7: Next generation mathematical databases

Mathematics has a rich notion of data: it can be either numeric or 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 common resource, and should be maintained as such. Much of this proposal, and the 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 common resource, while fostering a virtuous circle of interoperability between these different types of data: a mathematician might implement an algorithm, to be run later on numerical data collected by a scientist.

Objective 8: Collaborative Research Environments that Transcend Domains

Wide dissemination of our VRE is critical to ensure sustainability and reduce duplication of effort between communities. This dissemination is not restricted to the traditional arena of conferences, journal papers and workshops, but should exploit the high bandwidth communication provided by the internet. To ensure applicability of our framework, we will create a number of *demonstrators* to highlight the power of OpenDreamKit (**T4.6**, **T2.9**) 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* **T2.6** and books 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. Having been incorporated and developed by this project, they can be re-executed to serve both as a regression test and to form part of the documentation of OpenDreamKit.

Objective 9: Training and dissemination

The success of any research software or service is strongly linked to its ability to attract and retain a large number of users. The different communities (Sage, Gap, `PARI/GP`, Singular, `Jupyter`, ...) have each developed sustainable networks. For example, Sage has accumulated thousands of users in under 10 years. This has been achieved thanks to a very strong community building philosophy, especially through the organization of "Sage-Days" all over the world. The first Sage-Days was held in 2006; to date there have been at least 63 of them, including 10 during 2014, as well as Sage Education days, Sage Bug days, Sage Doc days, Sage Days aimed specifically at women, and more. Many of the OpenDreamKit project members have been involved in these events either as organizers or participants, and are convinced that they are a most efficient way to promote our software. More precisely, our objective is to create a constant dialogue between the different communities, through frequent workshops, conferences, user groups, and 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 minimising technical (non-research) obstacles to access existing tools: building better documentation and tutorials, developing easy-to-install distributions, enabling easy web and cloud access, better user interfaces, better interactions between different software. We will run a series of workshops to inject additional momentum into the process. By doing this, our objective will be 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 (WP4), and identifying, sharing and promoting software development best practices (Objective 3, 6 and 9).
VRE should integrate resources across all layers of the e-infrastructure (networking, computing, data, software, user interfaces)	OpenDreamKit will integrate resources across all layers of the e-infrastructure: software development models, collaborative tools, data, component architecture, deployment frameworks, standardization, social aspects (Objectives 1, 5 and 6, WP7), but also fostering collaboration inside the community, community enlargement and links with other scientific communities (Objectives 3 and 8, WP2 and WP4).
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 (Objectives 3, 4 and 5).
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 (Objectives 2, 3, 7 and 9), and make software sustainable, reusable and easily accessible (WP3 and WP6).
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 VRE, each tailored to a specific need in pure mathematics and application (Objectives 2 and 5). We will develop and demonstrate (WP2 and 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 requirements from use cases (WP2), including those involving industrial stakeholders. At the end of the project, the effectiveness of the VRE will be demonstrated for a number of use cases from different domains (Objectives 8 and 9).
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 `Jupyter` (`IPython`), connected to databases, that will allow a huge gain in efficiency and productivity, enabling a large-scale collaboration on software, knowledge, and data (Objectives 3, 4, 5 and 9, WP3 and WP5).
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 (Objectives 2 and 7, WP3 and WP6).

1.3 Concept and Approach

In its briefest form, the concept of this project is to develop, evaluate and disseminate a toolkit of compatible, open, modern and powerful software components from which bespoke VREs can be assembled to meet the needs of research projects in mathematics and its applications and developed and maintained by a sustainable free software ecosystem.

The next sections explain the history and role of computation in mathematics, and why our proposed toolkit will meet a wide range of research needs. We then explore the types of software and software development model which already exist, in order to explain the remaining tasks needed to realise our goal, and to maximise impact.

1.3.1 Background: Mathematics and innovation in the digital world

Mathematics is at the heart of innovation

We live in an innovation-driven society and mathematics is a key enabling tool for many of those innovations. The global positioning system (GPS) needs detailed calculations from special and general relativity. Computer Assisted Tomography (CAT scanning), a vital medical tool, is based on solving mathematical inverse problems. Mobile phone connectivity depends on combinatorial optimization algorithms and Delaunay triangulations for frequency allocation. The modern e-commerce infrastructure relies on cryptographic algorithms derived from number theory and algebra. At the core of each of these innovations there is underpinning mathematics developed and implemented as practical algorithms. These developments have been made possible through investments into pure and applied research in mathematics over many decades. Engineering and business innovation then builds on the mathematical insights to enrich society, though often the general public remains unaware of their mathematical foundations.

The mathematics research community has always been keen to develop and adopt new technology, from Newton's innovations in reflecting telescopes, to Turing and von Neumann's roles as the founding fathers of Computer Science. The power and adaptability 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.*".

This shows both the power of mathematical abstraction, that allowed two very different problems to be united, and the role of mathematical research as an "early adopter" of computational methods.

Even in more practical areas, such as in web standards, mathematicians have led 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. Computer algebra systems adopted and explored advanced programming concepts like comprehensions, iterators, or generics long before they became standard features of modern programming languages (e.g. in the early seventies for generics in Axiom, versus 2004 for Java).

Digital Exploration tools are crucial for research in mathematics From their earliest days, electronic computers have been used in pure mathematics, either to make tables, to prove theorems (famously the four colour theorem) or, as with the astronomer's telescope, to explore new theories. Computer aided experiments are now part of the standard toolbox of the pure mathematician, and certain areas of mathematics completely depend on it.

Computational experiments have led to new conjectures which have had a deep impact on the future development of mathematics. An outstanding example is the Birch and Swinnerton-Dyer conjecture (one of the Clay Millenium Problems). Databases relying on computer calculations such as the Small Groups Library in GAP, the Modular Atlas in group and representation theory, or the `LMFDB`, 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.

Computers as a tool for collaboration not just computation Mathematicians have always collaborated openly, and emphasised the role of the team of authors in a discovery. It is usual to list authors alphabetically on a paper, rather than the first or last author getting special credit, and it is normal to write about what "we" discovered, rather than assign credit to individuals.

In the last three decades, however, the mechanisms of collaboration have changed, and this is enabling a much more widespread and fine-grained collaboration. Whereas mathematics research on a day-to-day basis was traditionally a solitary pen-and-paper activity of talented individuals corresponding via lectures, letters, and journal articles, it is often now a collaborative, geographically distributed team activity supported by e-infrastructures.

E-mail was the first step in this direction, allowing correspondence in seconds instead of days, followed by web based tools such as the `arxiv` preprint server, bulletin boards like `mathoverflow.net` and collaboration tools such as `SageMathCloud` (see Section 1.3.9, page 17), Google Docs and `github`.

Collaboration on Mathematical Software, Data, Knowledge

Today research in many subjects is transformed by the availability of vast amounts of research data on the Internet. In this area mathematics has perhaps lagged behind, especially so when it comes to open exchange of data, not just static publication. Mathematical data is very varied and often has a complex structure. It could be divided into three kinds:

- *$\mathcal{D}$: tables or lists of numerical or symbolic data*
- *$\mathcal{K}$: 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 2 on page 12 shows some relationships between existing mathematical resources of these kinds.
Relevant examples include:

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* [http://www.lmfdb.org] 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. FindStat [http://www.findstat.org/] is an online database for statistics and maps on combinatorial objects. Its purpose is to automatically find relations between mathematical objects. It was initiated in 2011 and contains 228 statistics over 17 classes of objects.
   d. The libraries of small groups and semigroups (in GAP) which make accessible complete classifications of key algebraic objects up to a certain size.

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. The GAP library is roughly 400000 lines of code in a specially developed high-level language that describes many algorithms for diverse computations in algebra and discrete mathematics, not all of which are published elsewhere.
   b. A constraint solver such as Minion is by contrast a highly refined solution to a single problem (combinatorial search).
   c. The superseeker software provides an enhanced query interface to the Encyclopedia of Integer Sequences, detecting when the search key is a transformed version of a sequence in the database.

Many mathematical databases now exist, some very large; however their internal structure often hides the true richness of the data, limiting the scope for interaction with it to the specific tasks the designed had in mind. The past has shown that a more flexible 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 coincidence between dimensions of representations of the Monster group and coefficients of the j-function, and its conclusion eventually led to Borcherds' Fields medal;

Computational mathematics is interdisciplinary by nature The very name of "computational mathematics" suggests that it is interdisciplinary, drawing on both mathematics and computer science (and in fact having many applications in physics, chemistry and engineering as well). Computational techniques also open up unexpected connections between different areas of mathematics. For instance 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.

The diversity of needs in the mathematical community Certain scientific areas, for example in genomics, have large communities of researchers whose computational workflows are very standardized, which justifies the development of specialized Virtual Research Environments, typically taking the form of clickable web services. The situation is very different in mathematics.

Indeed mathematical research projects and teams that make use of computation, databases or collaborative tools are extremely diverse in size, skills, sophistication, needs, requirements, and available resources. Here are some typical scenarios to illustrate this:

- At one extreme a project might consist of a single researcher, with limited general computing expertise, using a computational tool to compute some data that confirms or refutes a hypothesis, writing up the results as a paper linking to the data and publishing it. Such a user needs a simple system that supports the computational tool of their choice, logging and replay of the computation, automatic incorporation of key data in a mathematical document, data archival and subsequent citation. They will use very little computational resource, and probably have no means of paying for what they do use, and will have limited ability to install software.
- A slightly larger research project in, for instance, algebraic combinatorics might involve two or three researchers of varying computing skills. It will require tools from very different areas of mathematics: linear algebra, commutative and non-commutative algebra, symbolic manipulations, group theory, graph theory, language theory, and rewriting techniques, perhaps. The researchers will thus need to use simultaneously many computational components, and to implement a specialised library of code that combines them in novel ways. They often will have access to some local or remote computational resource (cloud or HPC server), and parallelization and distribution of the computations is essential to cope with combinatorial explosion. Last but not least, they will need to visualize the results of their computations, typically large graphs with complicated information attached to the nodes or edges, and may have access to wall-sized screens for this. They will advertise their results through lectures involving live demos. Early on, they will want to share their code and data with colleagues typically with little computing expertise, and eventually contribute it back to the community.
- A larger collaborative project might involve five or six researchers at two or three sites, developing a significant extension to a system such as `Sage` and using it to explore or catalogue examples of the mathematics of interest, and publishing multiple versions of their software and data, and a number of mathematical papers based upon that data. They need a much more sophisticated environment, including communication and collaboration tools, software development tools and so on.
- Yet another type of project would be a very large and open-ended collaborative framework such as polymath, but with the capability to attach computations, machine-checked proofs and other computational elements to the discussion and the collaboratively assembled proof.

These are just a few of the many forms of computational or collaborative mathematical project that we aim to support.

1.3.2 Key concept: a Virtual Research Environment Toolkit

The diversity of requirements for different projects described above make it inappropriate to seek to provide a one-size-fits-all Virtual Research Environment, not even for substantial subcommunities. Instead we propose:

> Toward an Open Digital Research Environments Toolkit for the Advancement of Mathematics.
>
> OpenDreamKit proposes to deliver a flexible **toolkit** that will make it easy for individuals and teams of researchers of any size to set up custom collaborative Virtual Research Environments tailored to their specific needs, resources and workflows, which will provide modern, flexible and reliable support for the entire life-cycle of computational and collaborative work in mathematical research, from initial exploration to proof, publication, archival, teaching and outreach. They will support mathematical computations and databases of all scales from tiny to huge.

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 `Jupyter` notebooks; existing compute servers, clusters and clouds; typesetting tools such as LaTeX and so on.

1.3.3 Benefit of this concept: a toolkit which can evolve to best support real research practices

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.

Such engineering of 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 such an experiment has been a massive effort: `SageMathCloud` required 70k lines of bespoke code; similarly each of the databases and collaboration sites is essentially a bespoke program. Furthermore, much of this effort is necessarily *not* spent on the innovation in the environment, but on the underlying infrastructure. Our more flexible and modular system will allow focus to be placed on the environment itself. This ability to construct experimental VREs and gain feedback on them will accelerate the development of our understanding of the social dynamics of user and developer communities enormously, ultimately leading to better VREs.

1.3.4 Background: the scope of a VRE in mathematics, relevant recent developments

A unifying user interface: executable notebooks and project Jupyter Project `Jupyter` [31] is a set of open-source software projects for interactive and exploratory computing emerging from `IPython` [14]. These software projects help make scientific computing and data science reproducible and multi-language (Python, Julia, R, Haskell, Bash, R, . . .). The main component 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 (re-executable) code, equations, narrative text, interactive dashboards and other rich media.

Figure 1 shows a Python-based sample session. Within the Python session, all libraries available in Python can be imported and combined flexibly, a number of interfaces between different languages exist. Many more examples are available, for example [16] and within [38].

The `Jupyter` notebook is being used widely in academia (e.g. 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) as well as industry (e.g. Google, IBM, Facebook, Oracle, Otto Group, Microsoft, Bloomberg, JP Morgan, WhatsApp, O'Reilly, Quantopian, Logilab, GraphLab, Enthought, Continuum, Authorea, BuzzFeed) and journalism (e.g. 538 and New York Times).
Because the architecture and building blocks of `Jupyter` are open, they are 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.

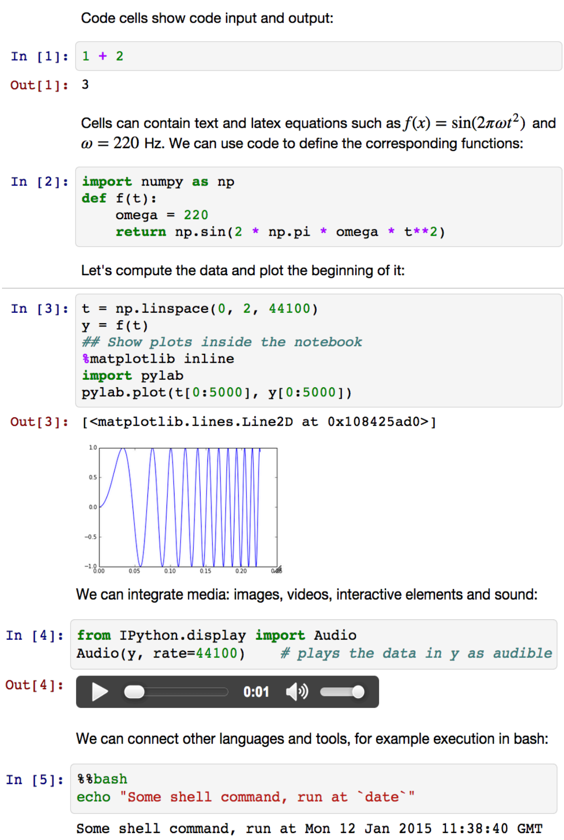

Figure 1: Self-contained `Jupyter` Notebook demonstrating the concepts of cells that contain different types of material and can be executed (or updated) in arbitrary or sequential order.

These notebook documents provide a *complete* and *executable* record of a computation that can be shared with others in a way that has not been possible before. This has led, among other things, to a huge boost in reproducible, interactive teaching/education documents in recent years. A paradigm that Fernando Perez, creator of the project, has referred to as "literate computing".[4]

We will build on this technology by extending `Jupyter` with new functionality, unifying other computational tools to be usable as components in this framework, and merging the `Sage` and `Jupyter` development.

[4] http://blog.fperez.org/2013/04/literate-computing-and-computational.html

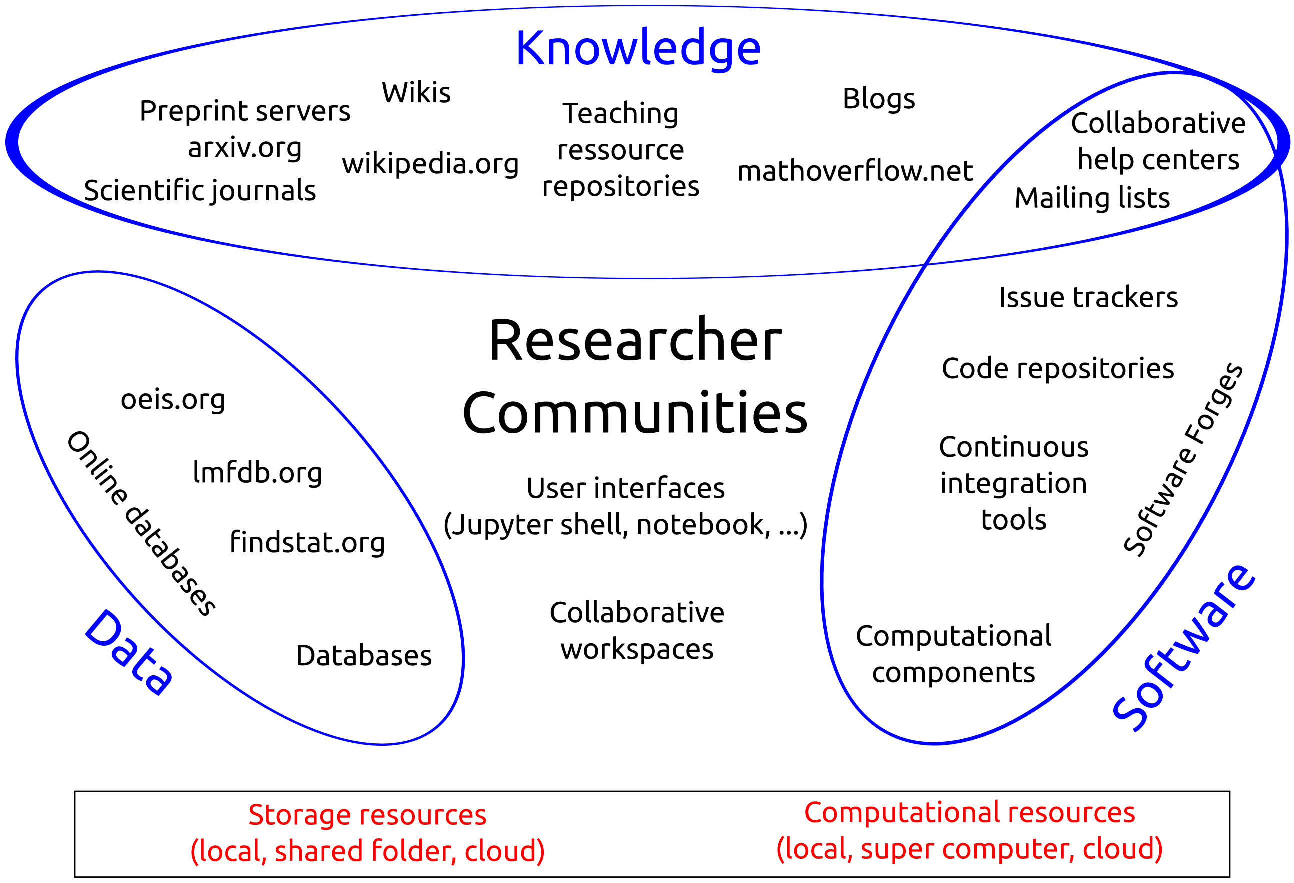

Figure 2: Virtual Research Environments for research in pure mathematics and applications.

Virtual Research Environments for Mathematics

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` (see Section 1.3.9, page 17) hosting more than 10k users and 100k projects after just one year. `SageMathCloud` is a open-source web-based hosting and web browser-based UI solution for full access to systems such as `Sage`, GAP, Singular, PARI/GP, IPython, and many more, to the end user, and may be descibed as truly *cloud*-based. Figure 3 on page 13 shows an example of an SMC session.

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 Finite 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.

Elements of a mathematical VRE can also be seen in “everyday” collaboration tools such as `arxiv` (sharing new mathematical knowledge with control of provenance) `Wikipedia` (presenting established knowledge in a consistent and linked way) and `github`, used for collaborative paper writing as well as software development.

Development models for mathematical software: a historical perspective Supporting the experimental method in mathematics requires spending substantial resources on software development. As the sophistication of the required computations increased, supported by the growth of available computational power, it became vital to distribute those efforts across ever-larger research communities. European mathematicians have been pioneers in this and have built up a tradition of collaborative open source software development, that was key to 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 much other scientific computing research relied on bespoke Fortran programs and used for one calculation and then discarded.

In that period, inter-project communication was limited by the lack of standardisation of, and interconnection between, computers. Systems produced remained limited to specific research topics and often specific computer systems and were not easily interoperable. It was left to the corporate world to gather sufficient manpower to develop general purpose systems

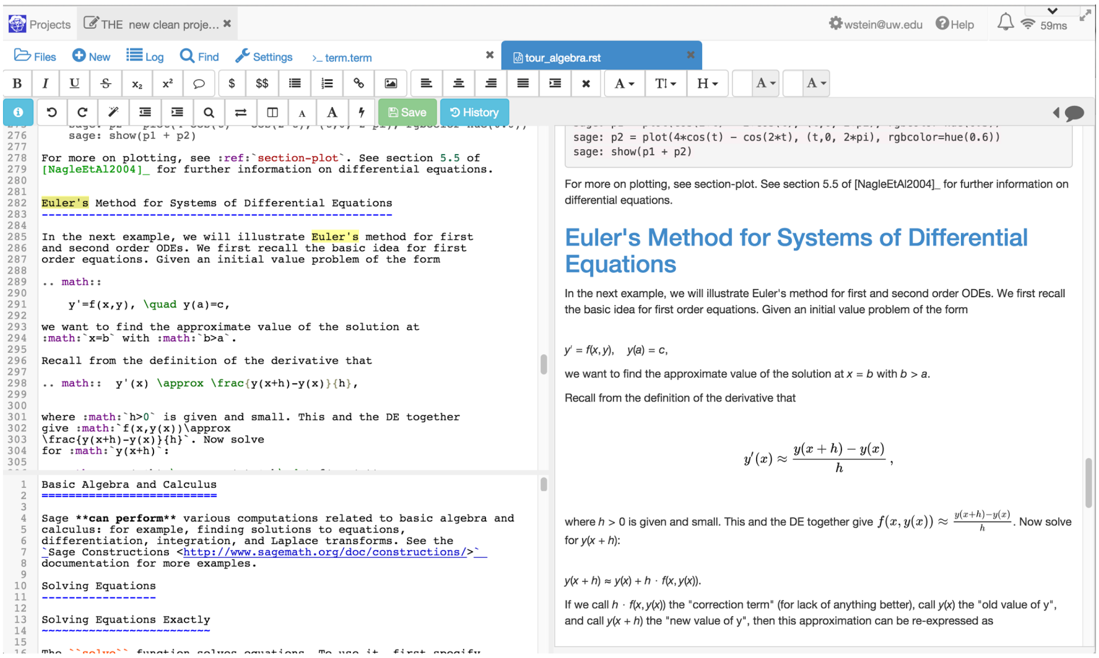

Figure 3: Typical `SageMathCloud` session in a web browser

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 open source software and open development models (the so-called “bazaar” approach). This is clearly exemplified in the `Sage` system. `Sage` is a truly general purpose computational mathematical system. It is committed to, and draws huge benefits from the power of open source software. 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` demonstrates this modern 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` (formerly 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) start to reverse the balance of these economies. Commercial packages can no longer develop fast enough to assimilate the innovation of the wider mathematical community, where there is greater expertise and manpower.

1.3.5 Approach

In previous sections, we have analyzed the diverse needs of researchers in pure mathematics and applications, and argued that the concept of a VRE toolkit, as proposed by OpenDreamKit, will match those needs, and have a considerable impact on how mathematical research is conducted.

In this section we describe how we will realize this concept by building on existing software components in a new way, and why this is an ideal time to do so.

The fundamental factor is that the last decade has witnessed the emergence of the necessary building blocks, all in open source:

- Key enabling technologies: virtual machines and containers for easy deployment; open cloud infrastructure; web technologies such as MathJax or WebGL allowing powerful in-browser clients; scalable decentralized database software and more
- Computational mathematics components as described elsewhere in this proposal
- Flexible user interfaces and interactive computing environments

The emergence of these components is itself enabled by progress in the wider technological environment – cheap powerful computers; fast reliable networks; stable and advanced platforms such as JavaScript; and more importantly and uniquely the maturation of open source development models and collaborative tools (e.g. `github`) supporting them. This now makes it possible to bring together large and diverse communities of developers, and foster large ecosystems of interoperable components. We elaborate later in this section how this has specifically affected the development of mathematical software in the last decade, showcasing the sustainability of the "by users for users" development models even for general purpose mathematical computational components. These models are still not perfect on the largest scales, and we will address this, for our purposes within the project.

> Our technical approach is to join forces with the `Jupyter` project and focus on developing and improving building blocks that can be assembled and re-used flexibly to address the varied requirements of mathematics and the applications of mathematics in science and engineering, rather than creating one particular monolithic environment.

Activities The activities of the project are planned and structured to develop and promote 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.

The project is divided into seven work packages. Work Package WP1 covers project management and coordination as usual. Work Package WP2 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 (**T2.1**, **T2.5**);
- 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 (**T2.1**, **T2.5**, **T2.6**);
- strengthening of virtual research communities (**T2.3**,**T2.5**);
- spreading of good practices, consultancy and training courses to new users (**T2.6**, **T2.8**);
- exchange of personnel and training of staff (**T2.3**).

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; (WP4, WP5)
- integration of installations and infrastructures into virtual facilities (WP3, WP7);
- innovative solutions for data collection, management, curation and annotation (WP6);

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 computing resources already available to them.

Networking activities The mathematical software community, and the various overlapping open source software communities already have an excellent spirit of collaboration, and excellent cultures promoting open debate, constant improvement and common purpose. Our networking activities build on this in essentially two directions: within the project we aim to provide more opportunities for close collaboration and learning from one another, in visits, code sprints, and cross-community days; outside the project we aim to encourage more users of mathematical computation to join this "community-of-communities" through our open workshops, training events and demonstrator applications.

Joint Research Activities The joint research activities are defined by our vision for OpenDreamKit. WP3 defines the requirements for a component to be part of the kit, WP4, WP5 and WP6 address limitations of current components, or areas where they have become out of date and WP7 is key to our goal to produce systems that are both effective and sustainable.

By focusing on a toolkit rather than a monolithic VRE, and by concentrating the efforts on improving and unifying existing general purpose building blocks, OpenDreamKit will simultaneously maximize sustainability and broad impact. Although the primary target users are *researchers in mathematics*, the beneficiaries include users of components such as `Jupyter` in scientific computing, physics, chemistry, biology, engineering, medicine, earth sciences and geography, and teachers and practitioners in industry as well as researchers. Users of many of the component systems will also benefit

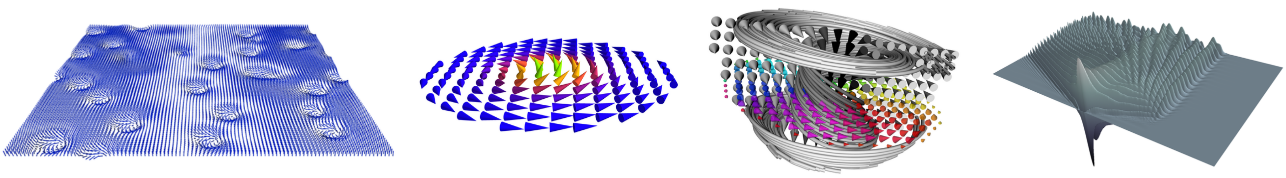

Figure 4: A selection of typical visualisation patterns often required in science and engineering. From left to right: a 3d vector field on a 2d domain, a 3d vector field coloured with another scalar field on a 2d domain, a 3d vectorfield on a 3d domain with streamlines, and scalar field plotted on a 2d domain.

from our improvements even if they do not choose to use OpenDreamKit VRE, since those components will be updated and improved. We will also perform important research into open development models that will benefit academia and many highly innovative SME's.

1.3.6 Benefits of this approach

Throughout this project we will reuse and extend existing open source systems. By doing this, we ensure that OpenDreamKit will benefit from future open source contributions during and beyond the lifetime 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. Finally, thanks to the "by users for users" model, the development will be steered by the actual needs of the community.

The net effect of this is that OpenDreamKit will include far more capabilities and code than could possibly be developed within a single research project, and will continue to accrue additional capabilities from its users as it is used. The investment in this project will be greatly amplified by these factors and the impact and sustainability greatly increased.

1.3.7 Demonstrators

Demonstrator: Training and education of the next generation computationalists — We will produce a series of interactive books that can be read, executed, modified and explored within the OpenDreamKit VREs, targetting Biology, Physics, Maths and Engineering students to demonstate the power of the approach and to train future scientists and engineers to embrace and exploit these advances in computational methodology (**T2.9**). We will also create additional materials for teachers at all levels to help them act as multipliers and disseminate the advances developed here widely (**T2.6**).

Demonstrator: Micromagnetic VRE — Micromagnetics is a continuum theory description of the behaviour of the magnetisation vector field at length scales of the order of micrometers and below. It is widely used in the research and development of magnetic data storage media and devices, for magnetic sensing, permanent magnets and healthcare applications such as cancer treatment and diagnostics. The mathematical model is a time dependent nonlinear partial differential equation with multiple length and time scales in the problem, and solution strategies are based on finite difference and finite element space discretisations and sophisticated numerical solution of the equations. As in many other applied research fields, the groups carrying out the simulations are often not the code developers, nor have they extensive computational background. More commonly, these are material scientists, engineers and physicists that use the simulation to interpret their experiments and support their device design planning. Industrial users include Seagate, Hitachi, TDK, Samsung, Bosch and Toyota.

Figure 4 shows magnetisation vector fields obtained in typical micromagnetic studies. They relate, from left to right, to a set of interacting magnetic skyrmions in a thin flim, a vortex in a thin Nickel film, a vortex in a half-sphere geometry, and the propagation of magnetic excitations in a 1d-system.

Here, we will use the OpenDreamKit components to put together a in micromagnetic VRE to (i) demonstrate the power of the approach in a concrete applied research setting, (ii) exploit that experience to evaluate the real value of the structure of this VRE and OpenDreamKit to a large and diverse set of end-users. In more detail, we will embed the most popular micromagnetic simulation software (Object Oriented MicroMagnetic Framework [1]) within a micromagnetic VRE, complement this with value-adding features, develop a number of executable documents inside this VRE that act as tutorials and documentation, disseminate the software and documents as open source and through workshops for the micromagnetic community.

We have chosen the OOMMF simulation package as the target tool because it is a somewhat typical representative of computational software: computation is driven through a text-based configuration file, data files are produced, and later processed, then figures are created from the processed data; leaving the scientist with the burden to link all these elements together. The benefits of moving to the integrated notebook workflow (see Section 1.3.4) are substantial. Furthermore, OOMMF is widely used (over 1800 recorded publications [2]) and thus provides benefit to an active and substantial community, who in return will provide plenty of feedback.

We evaluate the value of this demonstrator (**T7.4**), immediately feeding results back into the OpenDreamKit work. This will also be a case study for the sustainability of the approach and tool beyond the life time of this H2020 project.

1.3.8 Specific requests of the call

Use of Existing Basic Services This will be addressed primarily in work packages WP3, WP6 and WP5. Our architecture will include interfaces to standard APIs for cloud provisioning, authentication, cloud storage, HPC scheduling and so forth. Actual VREs will be deployed by users integrating those with specialist tools according to the needs of their projects.

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 the Head of Equality and Diversity at St Andrews, about any known gender differences in collaborative working and ensure that our collaborative tools properly support open, equitable and inclusive patterns of cooperation. This will be reported in deliverable D1.3.

Service discovery The OpenDreamKit web pages (**T2.1**) and the dedicated training portal (**T2.2**) will provide a central point of service discovery, providing a directory to all components, demonstrators, online services, protocol and interface specifications, software and other project activities and outputs.

1.3.9 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 WP3.

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. Sloan foundation that is supporting IPython development for two years (1/1/2013-12/31/2014), in particular at the University of California, Berkeley and California Polytechnic State University, San Luis Obispo. This grant enabled the project to focus on developing the IPython Notebook as a general tool for scientific and technical computing that is open, collaborative and reproducible. This goes a long way toward the aims 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.

Open Archive of Formalizations (DFG) The German Science Foundation (DFG) has funded a three-year project (OAF) on the development of a integrated knowledge base of formal mathematics. The project builds on the LATIN project (DFG) which developed a foundation-independent metalogical framework for representing and interrelating theorem prover logics based on the OMDoc/MMT format. In the OAF project, the LATIN Framework is extended to a system that allows combine the most important theorem prover logics in one system (MathHub/OAF). The PIs of the OAF project (Kohlhase/Rabe) are members of the OpenDreamKit Consortium.

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.

Cubicweb http://www.cubicweb.org

Logilab has been developing CubicWeb since 2001 as FLOSS (Free Libre Open Source Software). CubicWeb is a semantic web framework, that allows to build web applications and web services from an ontology. CubicWeb could be used in OpenDreamKit to build mathematical databases dynamically that will store data, knowledge and software.

Math Search (Leibniz Foundation) The Leibniz foundation has funded a three-year project on search in mathematical information systems. The project is undertaken by FIZ Karlsruhe (the publisher of the Zentralblatt Math Database) and JacU. The project has developed a formula/keyword search engine that is in production use for Zentralblatt (see https://zbmath.org/formulae/).

Simulagora http://www.simulagora.com

Logilab is maintaining Simulagora, a software as a service (SaaS) that builds on free software (FLOSS) to provide its users with a Virtual Research Environment (VRE) that greatly guarantees traceability and reproducibility as well as facilitates group collaboration. Logilab will bring its experience of Simulagora to OpenDreamKit and feed back to Simulagora many of the deliverables available under a free license.

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 partly at the University of Washington at Seattle, and there is a business plan for commercial support for courses.

In comparison OpenDreamKit focuses on open source building blocks and architecture to easily set up 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 be of benefit to SageMathCloud.

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. As well as providing a central repository of data as a resource for researchers in number theory, through its website www.lmfdb.org, the LMFDB provides a modern handbook of L-functions and related objects, including tables, formulas, links and references. The LMFDB has been funded by the NSF (2008-2012, $1M) and by EPSRC (2013-2019) through a £2.2M Programme Grant, PI Prof. J. E. Cremona (Warwick).

Almost all contributors to the LMFDB project are pure mathematicians, with good computational skills, but not professional programmers or software developers. The LMFDB hence needs to broaden the support it can call upon from software professionals, for the computation of number-theoretic data and also to support database management and enhance the website user interface. 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" (http://www.aimath.org/knowlepedia/), a very fruitful method of presenting mathematical knowledge which have been an unexpected spin-off, now used in many websites unrelated to the LMFDB.

The LMFDB project would benefit greatly from collaboration with OpenDreamKit by connecting with a wider pool of experts in computer science. The proposed joint workshops between the LMFDB and OpenDreamKit will stimulate and enable collaboration. 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 ICMS (Edinburgh) in 2013 on "Online databases: from L-functions to combinatorics", sponsored by the NSF, AIM and the ICMS.

"ACCORD: Algorithms for Making Complex Decisions on Structured Domains" is Edith Elkind's ERC Starter Grant awarded in 2014, and to be started in March 2015. It 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, and engineer positive outcomes by designing appropriate mechanisms. In particular, it aims to develop a suite of preference aggregation procedures with complex outputs (i.e., partial orders satisfying user-defined structural restrictions) that admit efficient algorithms on realistic inputs and are computationally resistant to dishonest behavior, and to identify a set of guiding principles that can be used to choose an appropriate procedure from this suite for a specific decision-making scenario.

OpenDreamKit VRE 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 grant, on in a "real life" situation, and the collaboration will be mutually beneficial for both projects.

"MathSoMac: The Social Machine of Mathematics", EPSRC EP/K040251/2 is Ursula Martin's EPSRC Senior Fellowship grant, started in 2014 and to be running for 4 years. It brings rigorous methods from social sciences into studying of the crowdsourcing, e.g., large-scale online collaboration, phenomenon in mathematical sciences. Most striking is this regard are recent large scale collaborations, such as the Polymath projects led by Fields medallists Gowers and Tao, the collaboration on homotopy type theory led by Fields medallist Voevedsky and others. Martin's project will develop new paradigms to understand these, and new tools to support them, within the framework provided by the larger EPSRC collaboration, SOCIAM, 2013-2018. SOCIAM aims to understand the phenomenon of social machines, defined as purposeful human interactions on the web, and to enable the effective co-ordination and deployment of the burgeoning ecosystem of social machines currently available. SOCIAM aims to answer questions such as how individuals are incentivised to take part, how communities develop and mature, and how the speed and quality of results can be optimised. A mathematical example is provided by OEIS, the online Encyclopeadia of Integer Sequences, where it is volunteer social mechanisms that determine the nature of the data available and the reliability and reproducibility of outcomes. SOCIAM in turn builds on previous work in Oxford (De Roure, Goble and others) on VREs, such as MyExperiment.

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

SCIEnce: Symbolic Computation Infrastructure for Europe (FP6 eRII3-CT-026133, 2006–2011) was coordinated by the University of St Andrews (PI Prof Steve Linton) and 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.

HPCGAP: High Performance Computational Algebra and Discrete Mathematics (EP/G055181, 2009–2014) was a 4-site project coordinated by the University of St Andrews (PI Prof Steve Linton). It 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.

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.

Doctoral Training Centre in Next Generation Computational Modelling http://ngcm.soton.ac.uk The Doctoral Training Centre in Next Generation Computational Modelling is a €12million investment, jointly funded by the UK's Engineering and Physical Sciences Research Council (EPSRC), the University of Southampton and 50 industry partners. Its mission is to improve professionalism, simulation software and exploitation of emerging hardware to support Computational Science and Engineering. The centre will train about 75 PhD students and is funded to run from 2013 to 2022.

The centre has chosen the `Jupyter` Notebook as the key tool used in its teaching programme, and runs a programme to improve and disseminate best-practice in computational science. The centre's PhD students will be natural contributors, testers, and target audience for dissemination of OpenDreamKit and its research results. Hans Fangohr (USO) is the director of this doctoral training centre.

1.4 Ambition

For most pure mathematicians using computational tools in their research, the state of the art at the start of 2015 still consists of a collection of separate programs, each of which must be installed individually on their desktop or laptop computer, respecting a complicated set of interdependencies. Alternatively, software may be installed for them on a departmental server or cluster, and used via a text- or browser-based remote login. The software performs computations (using a variety of excellent implementations of extremely sophisticated algorithms), with inputs and outputs usually in a bespoke text-based format. 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. The results of computations are incorporated into publications using cut-and-paste, and collaboration is mostly done through exchange of programs and data by email, shared general-purpose file servers or, rarely, a service such as GitHub. Amongst other problems, this approach creates a serious obstacle to the reproducibility of published computational experiments both by other researchers and the authors themselves at a later time.

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 many fields of pure mathematics, including for instance abstract algebra, number theory and algebraic geometry, and their design is often not well-suited to support these.

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 existing open 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, in particular `SageMathCloud`, `HPC-GAP`, SCIENCE (for all three, see 1.3.9); `Sage` and its notebooks; Polymath and MathOverflow (see MathSoMac entry in 1.3.9); and `Recomputation.org`. We will build on the experiences, and where useful, on 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.

The system will be **modern** in its construction: following best practices in distributed software development, internationalisation, use of web and clouds services, etc.; in its user experience, offering a modern supportive user interface that automates all of the routine tasks that it can; and in its support for important new research areas that may cross traditional subdiscipline boundaries. It will combine **seamlessly** a range of software components, hardware resources and databases, so that the user can work, or program with any combination of them in the same way (but, where relevant, can still attribute credit correctly). It will be **collaborative**, with shared projects the norm and discussion and exploration integrated with computation and writing. It will be **rigourous** in that, for instance, data passed between systems will be translated according to its mathematical meaning, not just its textual presentation. Finally it will be **adaptable** allowing an environment to be easily built and deployed to suit anything from a lone researcher tackling a problem for a week or two up to a complex project with subteams and multiple publications.

1.4.1 Challenges specific to mathematics

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

- The community mainly consists of individuals or *very* small groups (perhaps a professor and a few students). There are far fewer formal or structured research teams as you might find in an equipment-intensive science. There are certainly examples of large scale collaborations, for instance the project to prove the Classification of Finite Simple Groups in the 1980s and the Polymath experiments in the last few years, but these are driven by individuals, not defined by formal structures or funding bodies.
- Many top researchers have little or no formal research funding. If they need computational resources, these are limited to what is already available nearby, such as personal laptops or departmental clusters or to what they can access by asking favours of friends.
- Many mathematical computations are highly complex and irregular. Run times are not predictable and simple decomposition paradigms do not work well. Thus, traditional HPC approaches coming from numerical simulations and linear algebra do not apply.
- Mathematical notations have been refined over many centuries to be used 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 differ in several ways from typical scientific data
  - More often rather 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 search and reuse. Because of this, many issues (semantics, ontologies, reproducibility) involve the software which produced the data as much as the data itself.
  - The stored form of mathematical data (ultimately as strings of bits) is much further from the meaning of the data as perceived by a mathematician than is usual in other sciences. To make the link, many related objects and conventions must be considered and most interesting mathematical objects have multiple representations. Many mathematical theorems are implicit in these forms of representation, so that proving an ontology consistent may be very difficult.

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 SCIENCE 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.

Innovation Potential Nothing similar to the proposed OpenDreamKit VRE has been developed before, so the whole project is aimed at innovation. The closest model is `SageMathCloud`, the first usable VRE with extensive support specifically for pure mathematics.

It differs from OpenDreamKit in consisting of a single software component, deployed at a single site, and with no public API for other web services to build on it. Apart from a collaborative document editor it offers no support for publication of data or programs, or citability, or for automatic reproduction of published results. OpenDreamKit will make it easy for SMC and other VRE's build on this toolkit to address these and other limitations

The specific innovations in this project will also have have wider applicability. Indeed each and every improvement we will contribute to software components of the OpenDreamKit, and in particular key tools like `Jupyter`, will benefit their larger user communities (typically scientific computing) independently of whether they use VRE's or not.

2 Impact

The project, with its ambitions vision, general and broad approach, and challenging work plan, will offer the opportunity to all partners and beyond to complement their research expertise with methodologies and tools not available at their institutions. It will provide pivotal aspects needed for the development of a new generation of high-efficient scientific leaders with an open and constructive attitude toward collaborative interdisciplinary research and innovation. The diverse nature of the objectives composing the project will also be taken into account to design a successful and multiform dissemination and exploitation strategy.

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.

1. More effective collaborations between researchers	
OpenDreamKit will strengthen collaborations between European scientific community in different branches of mathematics, through a development of the common e-infrastructure by bringing together previously separated software and databases, and building 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 mathematical scientific community using the tool will increase by X%. 2. The scientific community in other disciplines (biology, physics, etc.) and industry using the tool will increase by X%. 3. The number of attendees of our training and dissemination events.
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	
The development of a new integrated tool, replacing previously separated tools, a possibility of real time data sharing, data re-use and simultaneous working will allow substantial gains in time and in efficiency, and, by consequence, higher productivity of researchers. Moreover, the exchange of best practice (such as regular code audit) will have an important impact on research quality. Finally, the unique tool will allow considerable reduction of further research costs.	Key performance indicators : 1. The e-infrastructure developed by the end of the project will allow to reduce the time of research by up to X% and to increase the productivity of researchers by up to X% 2. Due to the dissemination of best practices, research quality will considerably increase, and the number of errors will be reduced. 3. Better traceability and discoverability of research outputs. 4. Considerable reduction of research costs.

3. Accelerated innovation in research via an integrated access to digital research resources, tools and services across disciplines and user communities	
An integrated access to digital research resources and tools that OpenDreamKit will provide will clearly help to accelerate innovation in research across disciplines and communities. The integrated tool will meet the needs of all disciplines covered by the project and will help to overcome commonly found obstacles in both academic and non-academic research in these disciplines. Industrial actors, actively involved in the project, will directly benefit from the project results.	Key performance indicators : 1. Emergence of new, improved research methods and workflows in several disciplines using the tool. 2. Resolving of series of problems inherent to industrial actors in several disciplines. 3. The speed of resolution of issues across components (e.g. bugs reported by `Sage` users and problems encountered by `Sage` developers integrating `GAP` being resolved by the `GAP` team. 4. The number of components/libraries that are compatible with the OpenDreamKit user interface. 5. The number of components/tools/packages that may be used in parallel programs, or which incorporate parallelism in their own routines. 6. The number of joint publications and grant applications arising from the project 7. The number of external collaborators and contributors to OpenDreamKit.

4. Researchers able to process structured and qualitative data in virtual and/or ubiquitous workspaces	
OpenDreamKit will enable researchers to process different kind of data due to an integrated tool interconnected with databases. Efficient data storage will allow further exploitation and re-use of mathematical data for further calculations and thus make data processing more efficient.	Key performance indicators: 1. Number of data sets made available in mathematical databases supported by OpenDreamKit. 2. Number of OpenDreamKit components that are able to access those data sets, where relevant. 3. Number of remote queries to retrieve information from data sets hosted by project members.

5. Increased take-up of collaborative research and data sharing by new disciplines, research communities and institutions	
The project will clearly enhance a take-up of collaborative research and data sharing by and between new disciplines and research communities. Those communities already using parts of the tool, will be enlarged, involving more and more industrial actors and young scientists. By developing the tool, it will reinforce collaborations between different branches of mathematics (both pure and applied). Once developed, the tool will be opened for research in various disciplines, such as biology, physics, astronomy etc.	Key performance indicators : 1. Increase in collaboration and data sharing between different communities in mathematics. 2. Increase in collaboration and data sharing between research communities in other disciplines. 3. Increase in the number of users and research communities using the tool. 4. Increase in the number of publications citing OpenDreamKit and its components, and in the number of institutions and disciplines exhibited in these publications (although the timescales of publishing in some disciplines, including mathematics, mean that this will be very much a lagging indicator).

We will survey mathematical departments (and relevant members of other departments during year 1 and again during year 4 to gauge the awareness of the existence and capabilities of OpenDreamKit and its components, and to collect statistical data for estimating Key Performance Indicators from sections 2–5 of the table above. The success factor is a positive change between the two surveys. Other measures not listed in the table include activity on the relevant mailing lists and forums and the numbers of downloads (though this is hard to measure for many reasons).

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 (such as Avaya, Juniper, Extreme, Cisco, et al.);
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 (such as Accenture, HP, IBM, et al.)
6. End users: research communities; stakeholders in IT, healthcare, education, aeronautics, and other areas.

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 as well as to the needs of the scientific community. Moreover, this will allow short time-to-market and will facilitate the technology uptake.

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

Market needs	How the project will address these needs
Performance gain	The tool will enable its users to combine functionality from several major open-source mathematical software systems and problem-oriented programming languages, as well to the modern environment to use such mainstream programming language as Python on the majority of currently popular hardware platforms and operating systems.
Infrastructure capacity: newly built infrastructures with fast broadband connections are well positioned for adopting our products	OpenDreamKit will allow different groups of users to work simultaneously, and thus providing a considerable gain in efficiency.
Low scaling costs	An open source architecture brings affordability: people and organisations donate efforts towards common goals, and small organisations can gain access to equipment and research talents otherwise only affordable by the largest companies. Resources integration will reduce considerably the time and the costs of operations.
Going beyond limitations of interconnect technology	An open source architecture enhances creativity due to the potential to attract best minds to solve a problem
Enabling new applications and features	Through a series of connections 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 due to the new collaborative features. Liaising with industrial stakeholders during the development will allow to deliver a tool the suits their needs in the best possible way, thus speeding-up the time-to-market and technology uptake.
Easy-to-use service: first-time experiences are crucial to gain acceptance	We will design an ergonomic multi-user web-based graphical user interface, following web standards to best support a large array of browsers, including cell phones and tablets. We will explore opportunities for integration in interactive boards, as an aid for teaching and collaborative research.

2.1.3 Other impacts (environmental and socially important impacts)

Here, our mission statement will be based on the for `Sage`: “Creating a viable free open source alternative to `Magma`, `Maple`, `Mathematica` and `Matlab`”. We want to go beyond that goal, and to make OpenDreamKit a new reference tool, it is important to focus on the young generation, which will constitute its future users. This is known as “Generation Y phenomenon”: the so called “Generation Y” is expected to account for 30% of the total projected population in 2025, and would be the key influencer to change in workplace habits, caused by such features as easy adaptability to new technologies and social media, commonly attributed to this generation.

2.1.4 Potential barriers to impact

The following barriers to impact will be overcome using the mitigation strategies provided. These are distinct from the risks to project delivery detailed in Section 3.2.5. Table 2.1: Barriers to Impact

Barrier description	Risk level
Users will not use the new VRE environment	High
Contingency plan: A major concern in any proposal of this kind is that the resulting tools will not be adopted by users. This is a particular concern with such 'tradition-based' community as mathematicians. This project has two pathways to tackle this: (1) OpenDreamKit is based on prior work, which *already* has users; (2) OpenDreamKit will be integrated into the `Jupyter` and `Sage`, which *already* has a significant user base; (3) An end-user group formed at the beginning of the project by the representatives from different disciplines and sectors will provide a precious advice on real user needs and assist in providing OpenDreamKit sustainability. In addition, the project's communication, dissemination, and exploitation strategy will evolve throughout the project's implementation to ensure that stakeholder communities are fully aware of the project's progress, potential benefits, and innovative capacity and are engaged in the integration of the final results.	
Dominance of competing frameworks	Medium
Contingency plan: Our strategy is to engage with users and attract new users at the very beginning of the project, understand their requirements and design the domain-specific tools. An international advisory board will allow us to coordinate with the related research activities within and outside of Europe and to promote our framework internationally.	

2.2 Measures to Maximise Impact

The overall objectives of the dissemination and exploitation strategy are based on the project's core values, which are to improve the productivity of researchers in mathematics and connected fields by providing them with a unique virtual toolkit for a collaborative research tailored to their needs and requirements both during the project period and after the project completion.

2.2.1 Dissemination and Exploitation of Results

The strategy will be presented in the dissemination and exploitation plan, prepared by the Coordinator within the specifically designed WP and implemented with the help of all partners. The planned activities will bear in mind the project's scientific and societal impacts, and build throughout the project to ensure that stakeholder communities (1) are fully aware of the project and its potential benefits, (2) engaged in integration of the VRE in their professional activities, and (3) contribute to the sustainability and improvement of the VRE. In the following table, we summarise the dissemination activities and how they will help to achieve the expected impact among our stakeholders and target audiences.

Post-project Activities: The natural interest of the consortium is to ensure sustainability of OpenDreamKit also after the completion of the project. Therefore, the partners are committed to post-project efforts, which include the following activities:

1. Continue dissemination to scientific community and industrial stakeholders through participation to international conferences (FPSac, ISSAC, Python, `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. Gradual incorporation of OpenDreamKit components into the relevant university courses beyond OpenDreamKit members home institutions.
4. Expand OpenDreamKit user base by continuing the research collaboration with existing users and identifying new scientific (specifically from neighbouring fields) and industrial users.
5. Apply for funding at European / national levels for new projects that are to improve and further promote OpenDreamKit.

Exploitation: To exploit and capitalise on the success of the project, we will undertake the following activities

1. Engaging stakeholders and potential users of OpenDreamKit in Europe for realising the technology transfer and the innovation potential of the toolkit;

Dissemination goal	Target audience	Dissemination method	Timeframe and frequency
Project identity and profile	T1-T6	Website; flyers/leaflets; videos.	Throughout project, continuous
Broad dissemination	T3, T4, T6, T7	Biannual e-newsletters; press releases; information database; social networks and platforms; news in Nature and other editions in other disciplines; lectures in high schools led by PhD students.	Throughout project, quarterly
Knowledge transfer, information exchange	T1,T2, T3	Organisation of 10 technical workshops; 10 scientific trainings/year; training of at least 100 PhD students for the infrastructure usage; publications in social aspects; software demonstration during conferences; workshop for PhD students in Africa; participation at the workshop 'Sage and women' in US; participation in international conferences like FPSAC, ISSAC, or the international congress of mathematical software; regular participation in annual Python conference; organising at least 8 scientific trainings for other scientific communities/projects; news in Nature and other editions in other disciplines; certification by technology clusters.	Throughout project, continuous
Demonstration of advantages and possible applications	T1-T4	Organisation of 10 technical workshops; participation in annual Python conference.	Biannually
Uptake of the VRE by new users	T2-T4	White papers; organise at least 8 scientific trainings for other scientific communities/projects; presentation at international conferences; 5 MooCs designed to master students; integration of project results into Master courses and into teacher training courses.	Mo24-, at relevant milestones
Sustainable development beyond the project	T1-T4	Policy events; white papers; participation at conferences.	Mo24-, at relevant milestones

Key for the "Target Audience" column:

T1 Scientific community in mathematics and related fields (experienced researchers, under-/graduate/post-graduate students);
T2 Scientific community in other disciplines;
T3 Other relevant European and national initiatives and projects;
T4 Industrial end-users;
T5 Standardization agencies;
T6 Civil society;
T7 Public at large.

Table 2: Dissemination and exploitation plan

2. Teaching the OpenDreamKit-induced research results as parts of relevant courses at university, which are the home institutions of the consortium members, as well as at the on-going training events and activities (e.g., XXYY);
3. Mentoring and training PhD and postdoctoral students within the project as our contribution to educating excellent interdisciplinary European researchers in the strategically important domain of natural sciences;
4. Applying for European / other relevant funding programmes for new projects which arise due to the maturity of the OpenDreamKit;
5. Supporting spin-offs based on the developed technology by the consortium partners. In this case, the IPR management will be aligned with the corresponding rules set in the Consortium Agreement.

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.

By the end of the project, we expect that the main barriers will have been addressed, and that the needs in financial support after the end of the project will therefore not be very important. We further 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. In fact this opportunity is already being explored by the `SageMathCloud` project: it recently spun off a company on this business model to seek for additional funding.

Open access policy and data protection. OpenDreamKit will participate in the Open Research Data Pilot and is fully committed to ensure the open access of relevant project results and data. The consortium will comply with the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020.

The ambitious and interdisciplinary objectives of the project will result in the production of vast amount of research data (we refer to Figure 2 and the corresponding subsection for more details). The primary results of the project are expected to take the form of an open-source software, which will be available through the project website and publicly available repositories. Moreover, as the project strives towards efficient integration and representation of various research data and reproducibility of the research results, which represents naturally a challenge, the project will generate a detailed description of the data sources with specifics pertaining to data management (metadata standards, policies for access and sharing and for reuse and distribution, plans for archival and preservation, with accompanying deadlines). This information will be presented in the Data Management Plan, to be delivered within the first six months after the project start and subsequently updated throughout.

All scientific publications produced in the framework of the project will be either published in open access journals or self-archived using research data repositories. In addition, we will make all experimental data needed to reproduce/validate the results from scientific publications available through research data repositories (e.g. ZENODO, OpenAIRE).

Intellectual Property Rights Management. IPR management will be described in detail in the Consortium Agreement (CA), which will describe all issues regarding the IPR, confidentiality, know-how, rights on exploitation, the rights and obligations of the each partner. The CA will be prepared by the Coordinator, and then signed by all partners before the start of the project.

Access rights to foreground and background needed for the execution of the project shall be deemed granted, on a royalty-free basis, as of the date of the grant agreement entering into force. Methodology, documents, know-how, software, and tools will be available to all in order to achieve the project objectives during the project lifetime.

Most of the project results will have joint ownership due to a highly collaborative nature of the project. The CA will specify the terms of the resulting joint ownership, i.e., assignment of shares between joint owners, conditions of use, exploitation and management of jointly used IP.

The CA will also outline rules for publication procedures to ensure that IP can be protected while minimising publication delay.

The costs related to IPR (including those related to protecting results) and dissemination (i.e., 'gold' open access publications) are included in the project budget of each participating organisation.

2.2.2 Communication activities

Our intention is to increase the attractiveness of mathematics among young generation and females in particular as well as to improve the impact and maximise the visibility of the project activities on the entire VRE ecosystem. The following strategic access points will be used to maximise visibility:

1. An online presence that explains the OpenDreamKit concept and its applicability in layman's terms and offers significant information (website, social networks, Youtube, press releases).

2. Collaboration with other relevant European and national projects (existing and new ones). We refer to Section 1.3.9 for more details on the linked research and innovation activities. Presentation of the project results on the annual event 'Worldwide meetings of the free software.'
3. Collaboration with European and national mathematical societies, e.g., European Mathematical Society, European Women in Mathematics.
4. Presentations/demonstrations at partner institution-specific, locally organised 'science holiday' and 'days of science'.
5. Popularisation papers and communication events addressed to people interested by ICT.
6. Involvement in workshops / conferences on e-infrastructures and broad mathematical topics, e.g., Swiss Numerical Analysis Day.

3 Implementation

3.1 Work Plan — Work packages, deliverables and milestones

3.1.1 Overall Structure of the Work Plan

As shown in Figure 5, the work plan is broken down into seven work packages: WP3 about components, WP4 for user interfaces, WP5 for parallelisation of the components, WP6 for databases and finally WP7 for social aspects. This is complemented by the the usual management and dissemination work packages (WP1) and (WP2). The Gantt chart on Page 29 illustrates the timeline for the various tasks for these work packages

WP	Title	UPS	Logilab	UVSQ	UJF	CNRS	UO	USHEF	USO	USTAN	UW	JacU	UK	US	UZH	Simula	total
WP1	Management	*28*	2	2	2	2	2	2	2	2	2	2	2	2	1	2	**55**
WP2	Community Building/Dissemination	*10*	6	2		20		17	16	18			2	30		2	**123**
WP3	Component Architecture	64	14	*8*	6		4		6	16							**118**
WP4	User Interfaces	26	12	2		28		6	16	18		12	2	4		*28*	**154**
WP5	High Performance Math. Computing	6	12		*52*	40		12		18			60				**200**
WP6	Data/Knowledge/Software-Bases	4	2							10	25	*46*			12		**99**
WP7	Social Aspects						*23*	18	6								**47**
totals		**138**	**48**	**14**	**60**	**90**	**29**	**55**	**46**	**82**	**27**	**60**	**66**	**36**	**13**	**32**	**796**

Efforts in PM; WP lead efforts light gray italicised

Figure 5: Work Packages

3.1.2 How the Work Packages will Achieve the Project Objectives

The following table recalls the objectives of OpenDreamKit and lists the work packages that contribute to achieving each of them.

Objective	Purpose	WPs
Objective 1	Develop and standardize math soft and data for VRE	WP3, WP4, WP5, WP6
Objective 2	Develop core VRE components	WP3, WP4, WP5, WP6
Objective 3	Bring together communities	WP2, WP3
Objective 4	Update a range of softwares	WP3, WP5
Objective 5	Foster a sustainable ecosystem	WP3, WP4, WP5, WP6
Objective 6	Explore social aspects	WP7
Objective 7	Identify and extend ontologies	WP6
Objective 8	Effectiveness of the VRE	WP2, WP7
Objective 9	Effective dissemination	WP2, WP7

Work Programme for Objective 1: Objective 1 is covered by WP3 (component architecture) but also depends on WP4 (user interface) and WP5 (HPC).

Work Programme for Objective 2: Core components of the VRE are developed with work packages WP3, WP5 WP6 (databases) and WP4.

Work Programme for Objective 3: Diverse communities will be brought together within the work package WP3, especially with **T3.2** which aims to interface various softwares, and at the occasion of the community building and dissemination activities of work package WP2.

Work Programme for Objective 4: Individual pieces of software will be updated within work packages WP3 and WP5.

Work Programme for Objective 5: A sustainable ecosystem will be reached WP3, WP6, WP4, WP5 This objective should also benefit from WP7 (social aspects) with **T7.3**.

Work Programme for Objective 6: Objective 6 is covered by a dedicated work package WP7 on social aspects. It ranges from analysis of the needs with **T7.1** to evaluation with **T7.4**.

Work Programme for Objective 7: WP6

Work Programme for Objective 8: We will demonstrate and explore the effectiveness of the VREs through WP2 (dissemination) and the study of the social aspects of collaborative research in WP7.

Work Programme for Objective 9: The effective dissemination is achieved with part of WP2. More specifically, members of OpenDreamKit will organize workshops with **T2.5** or **T2.6**.

3.1.3 Work Plan Timing: GANTT Chart showing Task Dependencies and Information Flows

Since OpenDreamKit consists mainly in improving independent tools and integrating them into a VRE, its task are fairly independent from each other, which is reflected by the GANTT chart.

Figure 6: Overview Work Package Activities

3.1.4 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	UJF	DEM	PU	3
D5.2	Turn the Python prototypes for tree exploration into production code, integrate to `Sage`.	WP5	UPS	DEM	PU	3
D7.1	Report on relevant research in sociology of mathematics and lessons for design of OpenDreamKit VRE, part I	WP7	UO	R	PU	3
D2.1	Starting press release	WP2	UPS	DEC	PU	6
D3.11	Understand and document `SageMathCloud` backend code.	WP3	UPS	R	PU	6
D3.1	One-click install `Sage` distribution for Windows with Cygwin 32bits	WP3	UVSQ	OTHER	PU	6
D3.7	Virtual images and containers	WP3	UVSQ	OTHER	PU	6
D4.1	Active/Structured Documents Requirements and existing Solutions	WP4	JacU	R	PU	6
D5.3	Parallelise the relation sieving component of the Quadratic Sieve	WP5	UK	DEM	PU	6
D3.2	Python Interface to micromagnetic OOMMF package completed	WP3	USO	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	Logilab	DEM	PU	9
D6.1	$\mathcal{DKS}$ base survey and Requirements Workshop Report	WP6	JacU	R	PU	9
D6.2	Full-text Search (Formulae + Keywords) over LaTeX-based Documents (e.g. the arXiv subset)	WP6	JacU	OTHER	PU	9
D1.3	Internal Progress Reports, including risk management and quality assurance plan	WP1	UPS	R	CO	12
D2.2	Community building: Impact of development workshops, year 1	WP2	UPS	R	PU	12
D2.3	Demonstrator: Problems in Physics with Sage v1	WP2	US	DEM	PU	12
D2.4	Review on emerging technologies	WP2	UPS	R	PU	12
D3.5	Support for the SCSCP interface protocol in all relevant components (`Sage`, `GAP`, etc.) distribution	WP3	USTAN	OTHER	PU	12
D4.10	Tools for collaborating on notebooks via version-control	WP4	Simula	OTHER	PU	12
D4.2	Distributed, Collaborative, Versioned Editing of Active Documents in MathHub.info	WP4	JacU	DEM	PU	12
D4.4	Full featured `Jupyter` interface for GAP, `PARI/GP`, Singular	WP4	UPS	OTHER	PU	12
D4.6	`Sage` notebook / `Jupyter` notebook convergence	WP4	UPS	DEM	PU	12
D5.5	Sun Grid Engine support for Project Jupyter Hub	WP5	USHEF	OTHER	PU	12
D5.6	Make `Pythran` typing better to improve error information.	WP5	UJF	DEM	PU	12
D5.8	Extend the existing assembly superoptimiser for AVX and upcoming Intel processor extensions for the `MPIR` library.	WP5	UK	DEM	PU	12
D6.3	initial $\mathcal{DKS}$ base Design	WP6	JacU	R	PU	12
D4.3	Basic `Jupyter` interface for GAP, `PARI/GP`, `Sage`, Singular	WP4	UPS	OTHER	PU	14
D4.7	Micromagnetic VRE software completed	WP4	USO	OTHER	PU	15
D6.4	Design of Triform (DKS) Theories (Specification/RNC Schema/Examples)	WP6	JacU	R	PU	15

#	Deliverable name	WP	Lead	Type	Level	Due
D6.5	Implementation of Triform Theories in the MMT API	WP6	JacU	OTHER	PU	15
D1.6	Innovation Management Plan	WP1	UPS	R	CO	18
D2.21	Short Course: A short course for lecturers on using OpenDreamKit for delivering mathematical education.	WP2	USHEF	DEC	PU	18
D2.21	Short Course: A short course for lecturers on using OpenDreamKit for delivering mathematical education.	WP2	USHEF	DEC	PU	18
D4.8	In-place computation in active documents (context/computation)	WP4	JacU	DEM	PU	18
D4.9	Facilities for running notebooks as verification tests	WP4	Simula	OTHER	PU	18
D5.10	Report on development of designs for the `GAP` developments – parallel library, interacts to standard infrastructure and `Cython`-like extensions	WP5	USTAN	R	PU	18
D5.9	Implement a parallel version of Block-Wiederman linear algebra over GF2 and the triple large prime variant.	WP5	UK	DEM	PU	18
D6.6	`LMFDB` deep modelling: Fragment Identification & Initial Model Design	WP6	UZH	R	PU	18
D7.2	The flow of code and patches in open source projects	WP7	UO	R	PU	18
D6.7	Heuristic Parser for the OEIS Import, Cross Validation of $\mathcal{DKS}$-Model	WP6	JacU	OTHER	PU	20
D4.11	Micromagnetic VRE tutorial and documentation notebooks	WP4	USO	DEM	PU	21
D2.10	Impact of dissemination and training activities for years 1 and 2	WP2	UPS	R	PU	24
D2.16	Demonstrator: Linear Algebra - interactive book	WP2	US	DEM	PU	24
D2.19	Community-curated indexing tool (open source)	WP2	UVSQ	P	PU	24
D2.20	Community-curated indexing service for OpenDreamKit	WP2	UVSQ	DEM	PU	24
D2.22	Course material on using OpenDreamKit in data science.	WP2	USHEF	DEC	PU	24
D2.6	Community building: Impact of development workshops, year 2	WP2	UPS	R	PU	24
D2.19	Community-curated indexing tool (open source)	WP2	UVSQ	P	PU	24
D2.20	Community-curated indexing service for OpenDreamKit	WP2	UVSQ	DEM	PU	24
D2.22	Course material on using OpenDreamKit in data science	WP2	USHEF	DEC	PU	24
D3.12	*Personal* `SageMathCloud`: single user version of `SageMathCloud` distributed with `Sage`.	WP3	UPS	OTHER	PU	24
D3.13	Integration between `SageMathCloud` and `Sage`'s TRAC server	WP3	UVSQ	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	UVSQ	OTHER	PU	24
D4.12	Notebook Import into MathHub.info (interactive display)	WP4	JacU	DEM	PU	24
D4.13	Demonstrator online portal available	WP4	USO	DEC	PU	24
D4.14	`Jupyter` extension for 3D visualisation	WP4	Simula	OTHER	PU	24
D4.17	Refactorization of `Sage`'s `Sphinx` documentation system	WP4	UPS	OTHER	PU	24
D4.5	Python/Cython bindings for `PARI/GP`	WP4	CNRS	OTHER	PU	24
D5.11	Library design and domain specific language exposing `LinBox` parallel features to `Sage`	WP5	UJF	R	PU	24

#	Deliverable name	WP	Lead	Type	Level	Due
D5.12	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	UK	DEM	PU	24
D5.14	Provide FAT binary support for all new x86_64 processors, build system maintenance and improvements to tuning, profiling and testing subsystems for the `MPIR` library.	WP5	UK	DEM	PU	24
D5.7	Explore the possibility to interface smoothly `Pythran`, `Cython` and `Cilk++`.	WP5	UPS	DEM	PU	24
D6.8	Conversion of existing and new Databases to unified interoperable System	WP6	UZH	DEC	PU	24
D6.9	`Python/Sage` Computational Foundation Module in OMDoc/MMT	WP6	JacU	OTHER	PU	24
D7.3	TRAC add-on to manage ticket prioritisation	WP7	UO	OTHER	PU	24
D7.4	Report on relevant research in sociology of mathematics and lessons for design of OpenDreamKit VRE, part II	WP7	UO	R	PU	24
D7.5	Demo: Mechanism for comments on posted Jupyter notebooks	WP7	USHEF	DEM	PU	24
D6.10	Full-text search (Formulae + Keywords) over Notebooks	WP6	JacU	OTHER	PU	27
D2.11	Demonstrator: Problems in Physics with Sage v2	WP2	US	DEM	PU	30
D6.11	Modelling and importing `FindStat` into $\mathcal{DKS}$ base	WP6	JacU	OTHER	PU	30
D6.12	Formula search in CAS programs and Software Modules	WP6	JacU	OTHER	PU	30
D2.12	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
D2.13	Demonstrator: Nonlinear Processes in Biology interactive book	WP2	US	DEM	PU	36
D2.14	Community building: Impact of development workshops, year 3	WP2	UPS	R	PU	36
D2.23	Demonstrator: Interactive lecture notes and marking systems based on OpenDreamKit.	WP2	USHEF	DEM	PU	36
D2.23	Demonstrator: Interactive lecture notes and marking systems based on OpenDreamKit.	WP2	USHEF	DEM	PU	36
D3.4	Continuous integration platform for multi-platform build/test.	WP3	UVSQ	DEM	PP	36
D3.6	Semantic-aware `Sage` interface to `GAP`.	WP3	UO	OTHER	PU	36
D4.15	Computational Fluid dynamics visualization in web notebook	WP4	Simula	OTHER	PU	36
D4.16	Exploratory support for live notebook collaboration	WP4	Simula	OTHER	PU	36
D4.18	Exploratory support for semantic-aware interactive widgets providing views on objects represented and or in databases	WP4	UPS	DEM	PU	36
D5.13	Refactor and Optimize the existing combinatorics `Sage` code using the new developed `Pythran` and `Cython` features.	WP5	CNRS	DEM	PU	36
D5.15	Parallelise the Singular sparse polynomial multiplication algorithms	WP5	UK	DEM	PU	36
D5.18	Exact linear algebra algorithms and implementations. Library maintenance and close integration in mathematical software for `LinBox` library	WP5	UJF	DEM	PU	36

#	Deliverable name	WP	Lead	Type	Level	Due
D6.13	Python/Sage Declarative Semantics in OMDoc/MMT	WP6	JacU	OTHER	PU	36
D6.14	LMFDB Algorithm Verification with respect to a Triformal Theory	WP6	JacU	OTHER	PU	36
D7.6	Demo: Jupyter Notebook Live Poster	WP7	USHEF	DEM	PU	36
D6.15	Shared persistent Memoization Library for Python/Sage	WP6	USTAN	OTHER	PU	42
D6.16	Search from Notebooks/Active Documents Interface	WP6	JacU	OTHER	PU	42
D7.7	Report on relevant research in sociology of mathematics and lessons for design of OpenDreamKit VRE, part III	WP7	UO	R	PU	42
D7.8	Review of new publication mechanisms, including evaluation of demonstrator projects	WP7	USHEF	R	PU	42
D7.9	Game-theoretic analysis of development practices in open-source VREs	WP7	UO	R	PU	42
D2.18	Micromagnetic VRE workshops delivered	WP2	USO	OTHER	PU	44
D2.25	Demonstrator: Repository of interactive Notebooks in Machine Learning and Computational Biology based on OpenDreamKit.	WP2	USHEF	DEM	PU	44
D1.6	Innovation Management Plan	WP1	UPS	R	CO	45
D2.17	Demonstrator: Problems in Physics with Sage v3	WP2	US	DEM	PU	47
D2.24	Demonstrator: Computational Mathematics for Engineering	WP2	USO	DEM	PU	47
D5.19	Implementations of the GAP developments, ready for release	WP5	USTAN	OTHER	PU	47
D2.26	Community building: Impact of development workshops, year 4	WP2	UPS	R	PU	48
D2.27	Impact of dissemination and training activities for years 3 and 4	WP2	UPS	R	PU	48
D2.28	Ending press release	WP2	UPS	DEC	PU	48
D3.10	HPC enabled Sage distribution	WP3	UJF	OTHER	PU	48
D3.9	Packaging for major Linux distributions	WP3	UVSQ	OTHER	PU	48
D5.16	Parallel versions of the Singular sparse polynomial division and GCD algorithms.	WP5	UK	DEM	PU	48
D5.17	Ongoing support of Intel, AMD, ARM, Sparc processors and new Operating System versions for the MPIR library.	WP5	UK	DEM	PU	48
D5.20	Implementations of exact linear algebra algorithms on distributed memory et heterogenous architectures: clusters and accelerators. Solving large linear systems over the rationals is the target application.	WP5	UJF	DEM	PU	48
D5.21	final report and evaluation of the GAP developments	WP5	USTAN	R	PU	48
D6.17	LMFDB Integration of Algorithms, Data and Presentation	WP6	JacU	R	PU	48
D7.10	Micromagnetic VRE environment evaluation report	WP7	USO	R	PU	48

3.1.5 Milestones

The work in the OpenDreamKit project is structured by eight milestones, which coincide with the corresponding biyearly project meetings (except the Kickoff meeting). Given the nature of the project, with a large number of essentially independent tasks, there is no need for milestones attached to specific collections of tasks or deliverables. Instead, given that the meetings are the main face-to-face interaction points in the project, it's suitable to schedule the milestones for these events, where they can be discussed in detail, tracking the progress in each work package through status reports on the tasks and deliverables.

We envision that this setup will give the project the vital coherence in spite of the broad mix of disciplinary backgrounds of the participants.

Milestone number	Milestone name	Related work packages	Est. date	Mean of verification
M1	Project Meeting 1	WP1,WP2,WP3,WP4,WP5,WP6,WP7	6	Meeting report
M2	Project Meeting 2	WP1,WP2,WP3,WP4,WP5,WP6,WP7	12	Meeting report
M3	Project Meeting 3	WP1,WP2,WP3,WP4,WP5,WP6,WP7	18	Meeting report
M4	Project Meeting 4	WP1,WP2,WP3,WP4,WP5,WP6,WP7	24	Meeting report
M5	Project Meeting 5	WP1,WP2,WP3,WP4,WP5,WP6,WP7	30	Meeting report
M6	Project Meeting 6	WP1,WP2,WP3,WP4,WP5,WP6,WP7	36	Meeting report
M7	Project Meeting 7	WP1,WP2,WP3,WP4,WP5,WP6,WP7	42	Meeting report
M8	Project Meeting 8	WP1,WP2,WP3,WP4,WP5,WP6,WP7	48	Meeting report

3.1.6 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	**all**
Effort	*28*	2	2	2	2	2	2	2	2	2	2	2	2	1	2	**55**

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 M0-M3 Sites: UPS (lead), Logilab, UVSQ, UJF, CNRS, UO, USHEF, USO, USTAN, UW, JacU, UK, US, UZH, Simula, UWS

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 M6-M48 Sites: UPS (lead), Logilab, UVSQ, UJF, CNRS, UO, USHEF, USO, USTAN, UW, JacU, UK, US, UZH, Simula, UWS

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 M6-M48 Sites: UPS (lead), Logilab, UVSQ, UJF, CNRS, UO, USHEF, USO, USTAN, UW, JacU, UK, US, UZH, Simula, UWS

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:** 0
Site	*UPS*	Logilab	UVSQ	CNRS	USHEF	USO	USTAN	UK	US	Simula	**all**
Effort	*10*	6	2	20	17	16	18	2	30	2	**123**

Objectives

The objective of this work package is to further develop the community at the European scale, foster cross team collaborations, spread the expertise, and engage the greater community to participate in the definition and refinement of the requirements, and the implementation and use of the produced solutions. This includes:

- 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;
- providing training for partners inside the project, the community engaging with contributions to the project, and end-users of OpenDreamKit
- reviewing emerging technologies,
- develop demonstrators,
- 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.

Some of the networking activities and internal training will come from short to long term visits between the participants, to collaborate on specific features. A typical such visit would bring together an `Jupyter` developer with a GAP developer for a couple of days to implement a first prototype of a notebook interface to GAP.

A number of demonstrators will be developed in the project and disseminated in different ways.

All the code, documents, test and build infrastructure produced as part of the project will be made available as open source.

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

T2.1 Dissemination and Communication activities M0-M48 Sites: UPS (lead), USTAN

This task comprises all forms of direct dissemination and public communication activities such as press releases, creation of the project web-site (D1.6) 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. We will use standard community building technology such as mailing lists, Wikis and Forums, to support dissemination and engagement of the community to support this. We will also generate press releases at opportune moments (D2.1, D2.28).

T2.2 Training and training portal M0-M1 Sites: UPS (lead)

Training is at the heart of this project: through our open source approach, networking activities, workshops, demonstrators (**T2.10**), interactive books (such as in **T4.13** and **T2.9**), and training for teachers and trainers (**T2.6**), we have firmly integrated the training aspect into the core of our project plan.

Each of these activities will create dedicated webpages hosting the material to make it accessible to a public as wide as possible — inline with our philosophy for software, we believe in the benefits of sharing the material and maximising the value of the financial investment into this project.

In this task, we create a central OpenDreamKit training portal that serves as an inclusive point of entry to explore available training materials. This will be hosted on the projects website (**T2.1**).

T2.3 Community Building: Development Workshops M0-M48 Sites: UPS (lead), CNRS, UK, Simula, USTAN, USHEF

We will organize development workshops all throughout the project. The aim of these workshops is to bring together developers from the different communities to design and implement some key aspects of OpenDreamKit such as user interface, and documentation and to ensure cross compatibility. 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 and to kick-start new challenges. It is also a way to work within the communities we're reaching in 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 component (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. The currently anticipated list of workshops includes:

- 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-Singular 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 **T5.1**.
- Two Data Science workshops which use OpenDreamKit to develop effective machine learning and data modelling practice organised by Sheffield.
- 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.

Yearly reports will be delivered on the impact of these workshops on the community (D2.2, D2.6, D2.14, D2.26).

T2.4 Reviewing emerging technologies M0-M48 Sites: UPS (lead), USTAN, USO, USHEF, US, UVSQ, CNRS, Simula

In this task, we will produce periodic reviews (D2.4) of emerging technologies and relevant developments elsewhere, and implications for our plans, taking into account input from the communities. 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.

T2.5 Dissemination: reaching towards users and fostering diversity M0-M48 Sites: UPS (lead), CNRS, USHEF, USTAN

As lead developers of OpenDreamKit, most of us consider themselves as both scientists and developers. We have experience in reducing the gap between those two worlds. We organize training, workshops such as Sage-days to promote our tools and bring more users and developers from the scientific world. On the other hand, we often attend and present to more development-oriented gatherings like PyCon and SciPy to exchange with engineers and foster collaborations.

The aim of this task is to exploit this winning strategy within OpenDreamKit. Three events should be organized in the spirit of Sage-days to gather and train more users and foster scientific development around OpenDreamKit. These conferences usually welcome around 50 participants and have a big impact on the scientific community. One of them will be at CIRM in Marseille, another one at ICMS in Edinburgh and a third one probably in Dagstuhl Germany. In the same spirit, we will also have training sessions organised within the universities (Orsay and Grenoble). We will also run a series of 4 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.

The under-representation of women in the scientific world is even more perceptible when we intersect science with software development. As we know we have many talented women in our community, and we will organize some events targeted at women in the spirit of the "Women in Sage" days that happened many times in the US already. We are planning to have two of them in Orsay and at least one in Oxford where "Women in CS" days already took place.

Apart from these different events, we will also be present at important events of both our scientific community (international mathematical conferences such as FPSAC for combinatorics) and the python / open-source software development community: PyCon, SciPy, EuroPython, etc. The material we develop for presentation at these events will be made publicly available.

T2.6 Introduce OpenDreamKit to Researchers and Teachers M6-M44 Sites: USHEF (lead), USO

In this task, we will develop and deliver materials that will introduce OpenDreamKit to potential users—both researchers and teachers. Our initial focus will be on teachers, but as the results from WP7 become available we will deploy them with researchers, both local to the University of Sheffield and across the wider computational biology and machine learning fields.

We will develop a 'taster' seminar (1-2 hours) and follow-up short course (1-2 days) on OpenDreamKit for researchers and lecturers in all disciplines D2.21. 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 (http://ml.dcs.shef.ac.uk/gpss/). The Summer School is now in its fourth edition (over 140 students educated). The Roadshows have taken place in Uganda, Colombia and in 2015 they are scheduled for Italy, Australia and Kenya. The Kenya school will be the first to have more of a 'data science' focus that we think will be particularly appropriate for dissemination of OpenDreamKit (D2.22).

These seminars and short-courses will also 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 (D2.23). 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.

Finally, our aim will be to introduce ideas from WP7 in our teaching materials. By the end of the project we will have produced a series of interactive notebook demonstrators D2.25 of OpenDreamKit with a particular focus on computational biology, data science and machine learning. These notebooks will expand the use of VREs in these domains, appealing to researchers used to the domains of Bioconductor and MATLAB. We will make use of live notebook posters (D7.6) and commenting systems (D7.5). These interactive notebooks will be provided in a public repository (D2.25).

T2.7 Open source dissemination of micromagnetic VRE M24-M28 Sites: USO (lead), Simula, USHEF, UPS

Tasks **T4.11** and **T4.13** provide the micromagnetic VRE demonstrator (1.3.7) built on top of OpenDreamKit. In this task, we set up of the infrastructure (D2.12) 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 **T3.8** and **T4.11**) routinely when the micromagnetic VRE code or underlying OOMMF core code changes, (ii) re-execute notebooks (from **T4.13**) and use them as regression tests (using the outcome of task **T4.3**), and (iii) re-build downloadable installation files and virtual machine images.

T2.8 Micromagnetic VRE dissemination workshops M14-M40 Sites: USO (lead)

We will run a series of workshops (D2.18) during the evenings of 4 major international meetings on magnetism research[5] to disseminate the micromagnetic virtual research environment (Sect. 1.3.7 and **T2.7**) 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 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.

T2.9 Demonstrator: Interactive books M0-M36 Sites: US (lead), USO

One of the important elements of VREs is a common flexible writing 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 typesetting quality of desktop publishing software like LaTeX.

Recently, a few approaches tried to bring both interactivity and the typographic features. The modestly tagged markup language DocOnce targets the problem of reusability of the document source code for producing traditional LaTeX-based printed books, IPython notebooks, Sphinx documents (with Sage cells), and many other formats. MathBook XML is a lightweight XML application for authors of scientific articles, textbooks and monographs extensively using Sage cells for interactive elements. The Sphinx documentation software has been successfully applied for creation of interactive books containing Sage cells. Additional interactivity is offered using the Runestone tools.

The technical aspects of format for interactive publications is a subject of the task "Structured documents" in **T4.6**. In this task we will demonstrate usability of the results of **T4.6** in creation of scientific textbooks. Three interactive

[5] Anticipated most significant international meetings in the appropriate time frame are 61st Conference on Magnetism and Magnetic Materials (MMM2016), October 31-November 4, 2016, New Orleans, Louisiana; 62nd Conference on Magnetism and Magnetic Materials (MMM 2017), 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 (MMM2019), 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 in the past at these conferences and were well attended.

books will be created:

- Nonlinear Processes in Biology (D2.13)
- Linear Algebra (D2.16)
- Computational Mathematics for Engineering (D2.24)
- Problems in Physics with Sage/Python (D2.3, D2.11, D2.17)

The choice of those particular topics has been made for the sake of maximal diversity. The "Nonlinear Processes in Biology" will heavily use numerical solution of ODEs and PDEs, the Linear Algebra book will be a classical mathematical textbook, while the next book is targetting the engineering community. The last example will focus mostly on collaborative editing and modularity of content which is produced using VRE technologies. We will demonstrate the power of symbolic algebra where appropriate throughout. All books are natively designed within the VRE, pushing the next generation of learning methodology out into multiple communities.

The main research aspect for this task will be to integrate modern computational tools in classical scientific topics and explore how the VRE environment can accelerate the development and produce electronic documents with significantly enhanced pedagogy and learning effectiveness.

In particular we will answer following questions:

- When is a fully interactive worksheet required and when is a textbook with executable code cells sufficient?
- How to assemble a classical monograph by reusing independently working building block of text and code?
- What are best tools and practices for using a single source for producing printed and electronic (interactive) textbooks?
- How to collaboratively write reusable course material?
- How can we facilitate automatic testing of all code examples, plots, etc?
- How can students can benefit from using VRE?

T2.10 Demonstrator: Computational mathematics resources indexing service M20-M25 Sites: UVSQ (lead), CNRS

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 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; and teaching institutions host or link their own collections of pedagogical resources.

These collections are usually incomplete, limited in scope, hard to search, and difficult to keep up-to-date. 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 (D2.19) powering such service, and to host a (free) community-curated index for OpenDreamKit related resources as a demonstrator (D2.20).

Deliverables:

D2.1 (Due: 6, Type: DEC, Dissem.: PU, Lead: UPS) *Starting press release*

D2.2 (Due: 12, Type: R, Dissem.: PU, Lead: UPS) *Community building: Impact of development workshops, year 1*

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

D2.4 (Due: 12, Type: R, Dissem.: PU, Lead: UPS) *Review on emerging technologies*

D2.5 (Due: 18, Type: DEC, Dissem.: PU, Lead: USHEF) *Short Course: A short course for lecturers on using OpenDreamKit for delivering mathematical education.*

D2.6 (Due: 24, Type: R, Dissem.: PU, Lead: UPS) *Community building: Impact of development workshops, year 2*

D2.7 (Due: 24, Type: P, Dissem.: PU, Lead: UVSQ) *Community-curated indexing tool (open source)*

D2.8 (Due: 24, Type: DEM, Dissem.: PU, Lead: UVSQ) *Community-curated indexing service for OpenDreamKit*

D2.9 (Due: 24, Type: DEC, Dissem.: PU, Lead: USHEF) *Course material on using OpenDreamKit in data science*

D2.10 (Due: 24, Type: R, Dissem.: PU, Lead: UPS) *Impact of dissemination and training activities for years 1 and 2*

D2.11 (Due: 30, Type: DEM, Dissem.: PU, Lead: US) *Demonstrator: Problems in Physics with Sage v2*
D2.12 (Due: 32, Type: DEC, Dissem.: PU, Lead: USO) *Micromagnetic VRE code and documents source online*

D2.13 (Due: 36, Type: DEM, Dissem.: PU, Lead: US) *Demonstrator: Nonlinear Processes in Biology interactive book*
D2.14 (Due: 36, Type: R, Dissem.: PU, Lead: UPS) *Community building: Impact of development workshops, year 3*
D2.15 (Due: 36, Type: DEM, Dissem.: PU, Lead: USHEF) *Demonstrator: Interactive lecture notes and marking systems based on OpenDreamKit.*
D2.16 (Due: 24, Type: DEM, Dissem.: PU, Lead: US) *Demonstrator: Linear Algebra - interactive book*
D2.17 (Due: 47, Type: DEM, Dissem.: PU, Lead: US) *Demonstrator: Problems in Physics with Sage v3*
D2.18 (Due: 44, Type: OTHER, Dissem.: PU, Lead: USO) *Micromagnetic VRE workshops delivered*
D2.19 (Due: 24, Type: P, Dissem.: PU, Lead: UVSQ) *Community-curated indexing tool (open source)*
D2.20 (Due: 24, Type: DEM, Dissem.: PU, Lead: UVSQ) *Community-curated indexing service for OpenDreamKit*

D2.21 (Due: 18, Type: DEC, Dissem.: PU, Lead: USHEF) *Short Course: A short course for lecturers on using OpenDreamKit for delivering mathematical education.*
D2.22 (Due: 24, Type: DEC, Dissem.: PU, Lead: USHEF) *Course material on using OpenDreamKit in data science.*
D2.23 (Due: 36, Type: DEM, Dissem.: PU, Lead: USHEF) *Demonstrator: Interactive lecture notes and marking systems based on OpenDreamKit.*
D2.24 (Due: 47, Type: DEM, Dissem.: PU, Lead: USO) *Demonstrator: Computational Mathematics for Engineering*
D2.25 (Due: 44, Type: DEM, Dissem.: PU, Lead: USHEF) *Demonstrator: Repository of interactive Notebooks in Machine Learning and Computational Biology based on OpenDreamKit.*
D2.26 (Due: 48, Type: R, Dissem.: PU, Lead: UPS) *Community building: Impact of development workshops, year 4*
D2.27 (Due: 48, Type: R, Dissem.: PU, Lead: UPS) *Impact of dissemination and training activities for years 3 and 4*
D2.28 (Due: 48, Type: DEC, Dissem.: PU, Lead: UPS) *Ending press release*

Work Package 3: Component Architecture								**Start:** 0
Site	UPS	Logilab	*UVSQ*	UJF	UO	USO	USTAN	**all**
Effort	64	14	*8*	6	4	6	16	**118**

Objectives

The objective of this work package is to develop and demonstrate a set of APIs that enable 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 (such as Cloud-based, local, and server environments).

Description

This Work 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 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 and other users need to know what algorithms the software is going to run to solve a given problem.

T3.1 Portability M0-M36 Sites: UVSQ (lead), UPS

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-users.

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 M0-M36 Sites: UPS (lead), UVSQ, UO, USTAN

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 M0-M48 Sites: UVSQ (lead), UPS, Logilab

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 [7] and virtualization [41] 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 other hand, 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 M0-M48 Sites: Logilab (lead)

To deliver every six month a new Simulagora VM image containing all 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 M36-M48 Sites: UJF (lead), USTAN

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 be carried out in 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 M0-M24 Sites: UPS (lead), UVSQ

From its inception in 2013, SageMathCloud (see Section 1.3.9, page 17) has quickly developed into 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 set up 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 M6-M24 Sites: UVSQ (lead), UPS, Logilab
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 M3-M9 Sites: USO (lead), USTAN
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 **T4.11** to build the micromagnetic VRE demonstrator (Sect. 1.3.7).
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 targeting 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 [10] interface [29]. 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: UVSQ) *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: UVSQ) *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: UVSQ) *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: UVSQ) *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: UVSQ) *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: UJF) *HPC enabled Sage distribution*

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

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

D3.13 (Due: 24, Type: OTHER, Dissem.: PU, Lead: UVSQ) *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	US	*Simula*	**all**
Effort	26	12	2	28	6	16	18	12	2	4	*28*	**154**

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` (formerly `IPython` notebook) provides a browser based approach to constructing executable documents which comprise of code (in multiple languages), mathematics, text, diagrams (see Section 1.3.4). The framework is an ideal portal through which VRE can be operated. 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 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 unify the notebook infrastructure used in `Sage` with `Jupyter`, push forward dynamic documentation exploration capabilities, and work towards concurrent multi-user editing of notebooks. We will also develop exemplar `Jupyter` notebooks for education and research (e.g. **T2.9**).

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 VRE (1.3.7) as a demonstrator.

T4.1 Uniform notebook interface for all interactive components M0-M36 Sites: UPS (lead), Simula, UK, USHEF, USO, Logilab, USTAN, UVSQ

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

One of our objectives is to ensure the sustainability of the project (Objective 5). The current `Sage` notebook interface was developed alongside that of `Jupyter`, but with slightly different goals. A notebook interface for `Sage` is a vital integrative component, and development was fast tracked to ensure its availability to allow the project to move forward. However, `Jupyter`, whilst it initially proceeded more slowly, has a larger developer base and has now caught up with the `Sage` notebook in terms of functionality. In line with Objective 5 `Sage` will now phase out its own notebook and switch focus to the `Jupyter` notebook, outsourcing this key but non disciplinary component.

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 component of this task.

T4.2 Notebook improvements for collaboration M0-M24 Sites: Simula (lead), UPS, USHEF, JacU, USO, Logilab

In this task, we will further improve tools for collaboration between authors of shared `Jupyter` notebooks and draw from the experience of collaboration as set in Simulagora, SageMathCloud, etc.

Version control tools, such as Git and Mercurial, have become an integral part of open and collaborative science and software. Version control tools allow proposed changes to be reviewed ('diffing') and resolve conflicts through combination of changes ('merging'). `Jupyter` notebook documents are stored in text files as JSON formatted data. This makes them well suited to tracking in version control, but the JSON structure can make diffing and merging difficult. We will deploy tools to provide better support for visual diffing and merging of Notebook documents. These tools will be integrated into existing version control workflows D4.10. The MathHub.info system already has a distributed Git-based versioning system, which can serve as an entry point here.

Given the interactive nature of `Jupyter` notebooks, live collaboration, where multiple authors work on the document simultaneously (like in Google Docs) is particularly desirable. However, there are particular challenges for collaborative editing of *executable* documents. The potential for *shared execution* adds both value and challenge to the live collaboration. Some attempts have been made to deal with live collaborative sessions (e.g. SageMathCloud, Colaboratory) but so far these have been outside the core Jupyter project. In this task we will explore different models of single-notebook collaboration, including shared or separate execution D4.10. We will consider not only indicating authorship, but which author triggered which execution, and explore other challenges. Various avenues for live session collaboration will be explored for integration into Jupyter itself D4.16.

T4.3 Reproducible Notebooks M12-M24 Sites: Simula (lead), UPS, USO

In this task, we will develop tools that allow re-execution 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. Reproducibility dictates that the notebooks should 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 component of 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 D4.9. 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 M0-M36 Sites: UPS (lead), Simula, UVSQ

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 (D4.17) 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 M0-M12 Sites: UPS (lead), Simula, USO, UVSQ, Logilab

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 challenge 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,
- D4.18 (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 M0-M24 Sites: JacU (lead), Simula, USHEF, Logilab

Jupyter notebooks consist of a sequence of cells that contain either text or a program (see Section 1.3.4). 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 exploiting 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 design and build a solution that meets the expectations (D4.1). The implementation will be achieved through an iterative process that incrementally improves existing software solutions, making them interoperable and synergistic. Results of this convergence will be reported in D4.8, D4.6 and D4.12 and used in **T2.9**.

T4.7 Active Documents Portal M12-M36 Sites: JacU (lead)

We will extend the existing `http://mathhub.info` system to a portal for interacting with active/structured documents (see **T4.6**) and releasing the portal initially for internal use in the OpenDreamKit and later for general use. `MathHub.info` already provides very basic sTeX editing and versioning. In OpenDreamKit we will extend it on the computational side based on the integrated format from **T4.6**. The resulting portal will be made available to the consortium as D4.2 and would be used for semantically enhanced code documentation and knowledge representation (see WP6).

T4.8 Visualization system for 3D data in web-notebook M0-M24 Sites: Simula (lead), US, UPS, USO

The Jupyter notebook provides an attractive environment for building user interfaces for research. However, 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, as shown in Figure 4. Experimentations in low dimensional topology and differential geometry also relies on good drawing capabilities (e.g. SnapPy or SageManifolds based on IPython and Sage). 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. However, 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.

In this task we will first investigate available technologies for fast in-browser visualization of the typical structures to be displayed (isosurfaces, streamlines, vector fields, cross sections, etc.). There are several existing solutions which could provide basis for further development. One of the best known, and also advanced is three.js which provides basis for 3D visualization in a web browser. Three.js is WebGL based, but also provides canvas based rendering for system which do not support WebGL. It has already been experimentally deployed in Sage Cell Server and SMC projects. Another promising technologies are visualization libraries using exclusively OpenGL. They can be deployed in browser based systems by using of the WebGL API (which is a restricted subset of the regular OpenGL API). This can be accomplished by visualization executed purely in GPU. VisPy and glumpy projects have found GPU-only solutions for common visualization objects (lines, arrows, markers, text, iso-lines, iso-surfaces, text, etc) where data does not exit the GPU. The VisPy project already offers an experimental interface with Jupyter notebook that could be extended to cope with our specifications. Through this tight collaboration with the authors, OpenDreamKit could benefit from both dedicated and state-of-the art visualization techniques.

The SnapPy and SageManifolds projects will be considered for deployment of tools we develop (see associated deliverable D4.14).

T4.9 Visualization of 3D fluid dynamics data in web-notebook M12-M36 Sites: Simula (lead), US, UPS, USO

We propose to let computational fluid dynamics (CFD) be a driving application for the development of 3D visualization in Jupyter notebooks (**T4.8**) since CFD is one of the most demanding cases of scientific visualization. The same time this task (with deliverable D4.15) will be demonstrator for (**T4.8**).

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 would be initialised 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 (see also WP5). 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.

T4.10 Common option system for various displays in Sage M0-M24 Sites: CNRS (lead)

Given a mathematical object, it often has various possible representations on a computer. From raw text to LaTeX, from simple 2d picture to a complicated 3d animation.

In this task, we provide a uniform option system for displaying object within Sage (raw text, LaTeX, tikz, matplotlib, jmol, tachyon, . . .). We will implement some of the missing display and will benefit of the work done in task **T4.9**.

T4.11 Case study: micromagnetic VRE built from OpenDreamKit M9-M15 Sites: USO (lead), Simula, USHEF

In this task, we use the OpenDreamKit architecture to assemble a virtual research environment software tailored for the large micromagnetic research community (see Section 1.3.7).

The micromagnetic VRE will be based on the Jupyter notebook, the Python interface to the micromagnetic simula-

tion library OOMMF (**T3.8**), and the additional features added to Jupyter in this work package.
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 (Figure 4), and similar features for scalar fields such as field components and energies for static and time dependent data (linking to **T4.9**). 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 [30] and providing a continuous path from scripting to GUI. 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).

T4.12 Python/Cython bindings for Pari M0-M24 Sites: CNRS (lead)

PARI is a C-library and GP is its standalone interpreter. Partial Python/Cython bindings are provided by Sage.
The task aims to develop an independent Python/Cython library that would provide bindings for PARI/GP and which would tightly be developed within the PARI/GP team.
The deliverable for this task is D4.5.

T4.13 Demonstrator: micromagnetic VRE notebooks M15-M21 Sites: USO (lead), Simula, UPS

The purpose of the micromagnetic VRE (**T4.11**) 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 (D4.11) within the micromagnetic VRE including (i) a new tutorial on computational micromagnetics with OOMMF, (ii) the complete documentation of the OOMMF-Py library (**T3.8**), and (iii) a set of typical micromagnetic case studies. The tutorial, in terms of content, will take guidance from the tutorial provided for Nmag [28] 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 (**T4.14**).

T4.14 Online portal for micromagnetic VRE demonstrator M21-M24 Sites: USO (lead), Simula, JacU

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 (D4.13) which provides the micromagnetic VRE for anonymous use. This service allows users to execute the demonstrator tutorial and documentation notebooks (**T4.13**) and run the calculations in real time on the web server, without having to install any software on their own machine. This web service will be based on Docker [7] virtualisation technology and we will make available the scripts to create VirtualBox [41] images, and Docker containers. The same virtual machine images can also be used for Cloud hosted computing services.
We request €6000 to purchase a machine to provide these services (shared memory, 64 cores, 128GB RAM, Solid-state drive (SDD) to make the system more responsive).

Deliverables:

D4.1 (Due: 6, Type: R, Dissem.: PU, Lead: JacU) *Active/Structured Documents Requirements and existing Solutions* Presenting sTeX and Jupyter to the consortium, comparing and evaluating as stepping stones.

D4.2 (Due: 12, Type: DEM, Dissem.: PU, Lead: JacU) *Distributed, Collaborative, Versioned Editing of Active Documents in MathHub.info*

D4.3 (Due: 14, Type: OTHER, Dissem.: PU, Lead: UPS) *Basic Jupyter interface for GAP, PARI/GP, Sage, Singular*

D4.4 (Due: 12, Type: OTHER, Dissem.: PU, Lead: UPS) *Full featured Jupyter interface for GAP, PARI/GP, Singular*

D4.5 (Due: 24, Type: OTHER, Dissem.: PU, Lead: CNRS) *Python/Cython bindings for PARI/GP*

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

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

D4.8 (Due: 18, Type: DEM, Dissem.: PU, Lead: JacU) *In-place computation in active documents (context/compu-*

tation)

D4.9 (Due: 18, Type: OTHER, Dissem.: PU, Lead: Simula) *Facilities for running notebooks as verification tests*

D4.10 (Due: 12, Type: OTHER, Dissem.: PU, Lead: Simula) *Tools for collaborating on notebooks via version-control*

D4.11 (Due: 21, Type: DEM, Dissem.: PU, Lead: USO) *Micromagnetic VRE tutorial and documentation notebooks*

D4.12 (Due: 24, Type: DEM, Dissem.: PU, Lead: JacU) *Notebook Import into MathHub.info (interactive display)*

D4.13 (Due: 24, Type: DEC, Dissem.: PU, Lead: USO) *Demonstrator online portal available*

D4.14 (Due: 24, Type: OTHER, Dissem.: PU, Lead: Simula) *Jupyter extension for 3D visualisation*

D4.15 (Due: 36, Type: OTHER, Dissem.: PU, Lead: Simula) *Computational Fluid dynamics visualization in web notebook*

D4.16 (Due: 36, Type: OTHER, Dissem.: PU, Lead: Simula) *Exploratory support for live notebook collaboration*

D4.17 (Due: 24, Type: OTHER, Dissem.: PU, Lead: UPS) *Refactorization of Sage's Sphinx documentation system*

D4.18 (Due: 36, Type: DEM, Dissem.: PU, Lead: UPS) *Exploratory support for semantic-aware interactive widgets providing views on objects represented and or in databases*

Work Package 5: High Performance Mathematical Computing								**Start:** 0
Site	UPS	Logilab	*UJF*	CNRS	USHEF	USTAN	UK	**all**
Effort	6	12	*52*	40	12	18	60	**200**

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** (`PARI`),**T5.2** (`GAP`), **T5.3** (`LinBox`), **T5.5**(`MPIR`) and **T5.4** (`Singular`).

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 **T5.7**, 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 M0-M48 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.

The deliverables for this task are D5.3 and D5.9.

T5.2 GAP M0-M48 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. Target skeletons include irregular parallel maps and folds; transitive closure operations, especially orbits of group actions and chain reduction (as used in Gaussian elimination). Data structure targets include distributed task queue, hash table and array structures.Specific algorithm targets include linear algebra over finite fields; randomised search algorithms in general and matrix group recognition in particular and algorithms for analysing the structure of groups given by a finite presentation
- interfaces between `GAP` and standard cloud and HPC infrastructure. This work will be based on **T3.5**. At the moment `GAP` is designed for interactive use or use as a local (lacking resource control or security) SCSCP server. Data is taken from local files or fixed URLs. We will develop interfaces that make it easy to run `GAP` jobs through standard batch queue environments; enable SCSCP servers to take advantage of widely available authentication and resource control frameworks in cloud and HPC settings, and access resources through standard discovery and allocation mechanisms.

- 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.

The deliverables for this task are D5.10, D5.19 and D5.21.

T5.3 Linbox M3-M12 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. This will be addressed in deliverable D5.11. More generally the second aspect, addressed in deliverable D5.18 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, in deliverable D5.20, addresses the specificities of distributed computing, with a close focus on communications and heterogeneous infrastructures.

T5.4 Singular M0-M48 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. Output are D5.12, D5.15 and D5.16.

T5.5 MPIR M0-M48 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.

Outputs are D5.8, D5.14 and D5.17.

T5.6 HPC infrastructure for combinatorics M0-M6 Sites: UPS (lead), CNRS

Several members of the projects are experts in combinatorics a field where Sage is clearly a world leader [5]. 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 **T5.7**, **T3.3**, and **T3.5**. In particular, we will provide a mixed C/Python implementation that will be integrated within Sage and replace most of the Sage-combinat code (deliverable D5.2 and D5.13).

In a second and more exploratory direction, some experiments [11] shows that the large tree exploration problem is very easily solved in C++ using the new Intel `Cilk++` [36, 37] technology (See for example: https://github.com/jfromentin/nsgtree and https://github.com/hivert/IVMPG). We would like to explore the possibility to interface smoothly `Pythran`, `Cython`, and `Cilk++` (deliverable D5.7).

T5.7 Pythran M0-M24 Sites: Logilab (lead), UJF

`Cython` (a fork of `Pyrex`) is an extended-`Python` to C compiler that has received significant contributions from `Sage` developers, and is 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. Output is 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. Output is 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.

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. Output is D5.6.

T5.8 Sun Grid Engine Integration in Project Jupyter Hub M0-M12 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 D5.5.

Deliverables:

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

D5.2 (Due: 3, Type: DEM, Dissem.: PU, Lead: UPS) *Turn the Python prototypes for tree exploration into production code, integrate to `Sage`.*

D5.3 (Due: 6, Type: DEM, Dissem.: PU, Lead: UK) *Parallelise the relation sieving component of the Quadratic Sieve*

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

D5.5 (Due: 12, Type: OTHER, Dissem.: PU, Lead: USHEF) *Sun Grid Engine support for Project Jupyter Hub*

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

D5.7 (Due: 24, Type: DEM, Dissem.: PU, Lead: UPS) *Explore the possibility to interface smoothly `Pythran`, `Cython` and `Cilk++`.*

D5.8 (Due: 12, Type: DEM, Dissem.: PU, Lead: UK) *Extend the existing assembly superoptimiser for AVX and upcoming Intel processor extensions for the `MPIR` library.*

D5.9 (Due: 18, Type: DEM, Dissem.: PU, Lead: UK) *Implement a parallel version of Block-Wiederman linear algebra over GF2 and the triple large prime variant.*

D5.10 (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*

D5.11 (Due: 24, Type: R, Dissem.: PU, Lead: UJF) *Library design and domain specific language exposing LinBox parallel features to Sage*

D5.12 (Due: 24, Type: DEM, Dissem.: PU, Lead: UK) *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*

D5.13 (Due: 36, Type: DEM, Dissem.: PU, Lead: CNRS) *Refactor and Optimize the existing combinatorics Sage code using the new developed Pythran and Cython features.*

D5.14 (Due: 24, Type: DEM, Dissem.: PU, Lead: UK) *Provide FAT binary support for all new x86_64 processors, build system maintenance and improvements to tuning, profiling and testing subsystems for the MPIR library.*

D5.15 (Due: 36, Type: DEM, Dissem.: PU, Lead: UK) *Parallelise the Singular sparse polynomial multiplication algorithms*

D5.16 (Due: 48, Type: DEM, Dissem.: PU, Lead: UK) *Parallel versions of the Singular sparse polynomial division and GCD algorithms.*

D5.17 (Due: 48, Type: DEM, Dissem.: PU, Lead: UK) *Ongoing support of Intel, AMD, ARM, Sparc processors and new Operating System versions for the MPIR library.*

D5.18 (Due: 36, Type: DEM, Dissem.: PU, Lead: UJF) *Exact linear algebra algorithms and implementations. Library maintenance and close integration in mathematical software for LinBox library*

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

D5.20 (Due: 48, Type: DEM, Dissem.: PU, Lead: UJF) *Implementations of exact linear algebra algorithms on distributed memory et heterogenous architectures: clusters and accelerators. Solving large linear systems over the rationals is the target application.*

D5.21 (Due: 48, Type: R, Dissem.: PU, Lead: USTAN) *final report and evaluation of the GAP developments*

Work Package 6: Data/Knowledge/Software-Bases							**Start:** 0
Site	UPS	Logilab	USTAN	UW	*JacU*	UZH	**all**
Effort	4	2	10	25	*46*	12	**99**

Objectives

The ultimate purpose of a mathematical VRE is to create *data* ($\mathcal{D}$; see Section 1.3.1), *knowledge* ($\mathcal{K}$), and *software* ($\mathcal{S}$) by modeling world situations, computing mathematical objects, and running computational experiments. To be effective a VRE needs an infrastructure that supports the creation, management, access, and dissemination of $\mathcal{DKS}$-Structures. All the systems considered in this proposal (GAP, Sage, PARI, Singular, OEIS, arXiv.org, ...) already include data, knowledge, and software modules as part of their regular distribution, but not in a form that is interoperable between systems, severely limiting the usefulness of the systems and results. The objectives of this work package are

1. to design metadata and representation formats for trans-system $\mathcal{DKS}$ structures as a basis for a math VRE,
2. implement interfaces to existing systems for interoperability and compatibility with the RE, and
3. implement a joint $\mathcal{DKS}$ infrastructure for, searching, documentation, traceability, versioning, provenance, visualisation and native dissemination of OpenDreamKit results (the latter three together with WP4).

Concretely we will 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.

User 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.

Description

We need ways to represent $\mathcal{DKS}$ in the same representational system, make the $\mathcal{DKS}$ structures explicit and therefore machine-manageable and – since current computational/experimental mathematics involve quite extensive $\mathcal{DKS}$– we need a new kind of "database", which we will call Mathematical Data/Knowledge/Software-base ($\mathcal{DKS}$ **base**), and which we will build in this work package.

The starting points for this unification effort will be the system-oriented data bases for $\mathcal{D}$, the OMDoc (Open Mathematical Documents) framework [18] for $\mathcal{K}$. OMDoc/MMT [33] is a representation format for mathematical documents and knowledge that incorporates a metalogical framework to be foundation-independent, which allows interoperability between various ontologies/foundations of mathematics. For the integration of $\mathcal{K}$ and $\mathcal{S}$ we will build on the notion of *biform theories* developed by Carette/Farmer [9] and extended to OMDoc/MMT by JacU in [20]. In this setup, the programming language serves as a foundation, just as ZFC set theory might for mathematical knowledge. Complex relationships between mathematical objects, interpretations of the underlying languages, and unit transformations are modeled in a graph of theories and theory morphisms.

The complexity of mathematical $\mathcal{DKS}$ structures 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.

T6.1 Survey of existing $\mathcal{DKS}$ bases, Formulation of requirements M0-M3 Sites: UZH (lead), JacU, USTAN, UW, US

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.

We will organise a workshop associated to this task (see **T2.3**). Results will be communicated in D6.1.

T6.2 Triform Theories in OMDoc/MMT M0-M12 Sites: JacU (lead), UZH

Work here would extend OMDoc/MMT biform theories along the data axis, which will require a specialised but integrated treatment. This integration will serve as a theoretical basis informing the design of a $\mathcal{DKS}$ base in **T6.3**. The results are reported in D6.4 (report) D6.5 (implementation).

T6.3 $\mathcal{DKS}$ Base Design M6-M12 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.

The initial design of the $\mathcal{DKS}$ base in OpenDreamKit is reported in D6.3. Conversion issues are covered in **T6.4**.

T6.4 Computational Foundation for Python/Sage M6-M18 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 the programming language 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);
- *specification*: what are the observable properties of the computation.

The foundation (a triform theory in OMDoc/MMT) will be published as D6.9.

T6.5 OEIS Case Study (Coverage and automated Import) M12-M18 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 [6]) 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 data is licensed under a Creative Commons 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 community resource.

The results of the import are reported in D6.7.

T6.6 FindStat Case Study (Triformal Theories) M18-M30 Sites: JacU (lead), UZH

In this task we would develop triformal theories for the `FindStat` project [7] 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 WP4, 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. Results will be reported in D6.11.

T6.7 `LMFDB` Case Study (Triformal Theories) M12-M24 Sites: JacU (lead), UZH, UW

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 (see D6.6).

Then we will perform cross-validation of the three model parts against each other (essentially model-based testing of software and inference; see D6.14). Once this has been successful for the chosen fragment, we will try to semi-automatically extend the import and model to the whole `LMFDB` to gain coverage and integrate it fully into the $\mathcal{DKS}$ base. We expect that this will entail quite a lot of work in refactoring the `LMFDB` proper, which will benefit the `LMFDB` community independently of its use in OpenDreamKit.

Finally, we will pick an algorithm from the `LMFDB` and verify it against its specification and the computational foundation developed in **T6.4**; this is the final validation of the case study. The results are reported in D6.17.

T6.8 Memoization and production of new data M24-M42 Sites: USTAN (lead), US, UPS, UW

Many CAS users run large and intensive computations, for which they want to collect the results while simultaneously working on software improvements. `GAP` retains computed attribute values of objects within a session; `Sage` currently has a limited `cached_method`. Neither offers storage that is persistent across sessions or supports publication of the result or sharing within a collaboration. We will use, extend and contribute back to, an appropriate established persistent memoization infrastructure, such as `python-joblib`, `redis-simple-cache` or `dogpile.cache`, adding features needed for storage and use of results in mathematical research. We will design something that is simple to deploy and configure, and makes it easy to share results in a controlled manner, but provides enough assurance to enable the user to rely on the data, give proper credit to the original computation and rerun the computation if they want to. Results are reported in D6.15.

T6.9 Math Search Engine M3-M9 Sites: JacU (lead)

The advantage of having a unified $\mathcal{DKS}$ base for a math VRE is that we can navigate the combined information space of all the underlying tools, systems and resources integrated into the VRE. The negative effect is that this aggravates the already serious problem of finding anything. Therefore we will adapt the existing MathWebSearch Engine (MWS [21, 26]) to the $\mathcal{DKS}$ base system. MWS consists of a web service that indexes mathematical

[6] `http://www.oeis.org`
[7] `http://www.findstat.org`

documents (formula/text) and a web front-end that allows users to query the index. Formula queries are highly efficient (25μs/query) and can be combined with keyword/full text search queries. An initial search engine for papers and system documentation will be established early in the project (see D6.2). For an integration into the $\mathcal{DKS}$ base we only need to build new harvesters – i.e. programs that generate keywords and formula URL/pairs from the contents (see D6.10).

For the data/software components in $\mathcal{DKS}$ this is true in principle, but the formulae in code can take many more forms and the notion of a hit URL is not as clear. But the theory graph structure and foundation change morphisms can be integrated into search so that even systems that are incompatible at first glance can be searched under one interface [19].

But this puts high demands on the search interface (user inputs are usually only meaningful with respect to a given context). We will explore this together with the notebook development – semantically annotated notebooks and active documents serve as an explicit context here together with WP4; results of this integration will be reported in D6.16.

Deliverables:

D6.1 (Due: 9, Type: R, Dissem.: PU, Lead: JacU) *$\mathcal{DKS}$ base survey and Requirements Workshop Report*

D6.2 (Due: 9, Type: OTHER, Dissem.: PU, Lead: JacU) *Full-text Search (Formulae + Keywords) over LaTeX-based Documents (e.g. the arXiv subset)*

D6.3 (Due: 12, Type: R, Dissem.: PU, Lead: JacU) *initial $\mathcal{DKS}$ base Design*

D6.4 (Due: 15, Type: R, Dissem.: PU, Lead: JacU) *Design of Triform (DKS) Theories (Specification/RNC Schema/Examples)*

D6.5 (Due: 15, Type: OTHER, Dissem.: PU, Lead: JacU) *Implementation of Triform Theories in the MMT API*

D6.6 (Due: 18, Type: R, Dissem.: PU, Lead: UZH) *LMFDB deep modelling: Fragment Identification & Initial Model Design*

D6.7 (Due: 20, Type: OTHER, Dissem.: PU, Lead: JacU) *Heuristic Parser for the OEIS Import, Cross Validation of $\mathcal{DKS}$-Model*

D6.8 (Due: 24, Type: DEC, Dissem.: PU, Lead: UZH) *Conversion of existing and new Databases to unified interoperable System*

D6.9 (Due: 24, Type: OTHER, Dissem.: PU, Lead: JacU) *Python/Sage Computational Foundation Module in OMDoc/MMT*

D6.10 (Due: 27, 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**

D6.11 (Due: 30, Type: OTHER, Dissem.: PU, Lead: JacU) *Modelling and importing FindStat into $\mathcal{DKS}$ base*

D6.12 (Due: 30, Type: OTHER, Dissem.: PU, Lead: JacU) *Formula search in CAS programs and Software Modules*

D6.13 (Due: 36, Type: OTHER, Dissem.: PU, Lead: JacU) *Python/Sage Declarative Semantics in OMDoc/MMT*

D6.14 (Due: 36, Type: OTHER, Dissem.: PU, Lead: JacU) *LMFDB Algorithm Verification with respect to a Triformal Theory*

D6.15 (Due: 42, Type: OTHER, Dissem.: PU, Lead: USTAN) *Shared persistent Memoization Library for Python/Sage*

D6.16 (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.

D6.17 (Due: 48, Type: R, Dissem.: PU, Lead: JacU) *LMFDB Integration of Algorithms, Data and Presentation*

Work Package 7: Social Aspects				Start: 0
Site	*UO*	USHEF	USO	**all**
Effort	*23*	18	6	**47**

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 itself is the ground truth. The traditional model of mathematical research is a mathematician, or a small group of mathematicians, standing around a blackboard, producing a proof they would "clean up": remove all traces of the process that led to its discovery and then submit the "clean" text to their peers for review.

Mathematicians have adopted new technology in a variety of ways: email and shared documents are used to collaborate on problem-solving and writing; larger "crowdsourcing" [34, 39], arrangements pull together diverse experts; symbolic computation tackles huge routine calculations; and computers check proofs that are too long and complicated for a human to comprehend. These technologies reveal (since email messages, version control systems and bulletin boards can be analysed) and alter the ways in which mathematicians collaborate.

In an EPSRC funded project "The Social Machine of Mathematics" Martin and others are bringing together rigorous methods from the social sciences to study these collaborative processes. Combining this research with the algorithmic game theory expertise of Elkind and Pasechnik, in this work package we intend to pursue the following objectives:

- incorporate the insights from this and similar projects into the design of OpenDreamKit VRE, ensuring that it supports the ways in which mathematicians really work, rather than the way software developers—or indeed mathematicians—think they do;
- extend this work to study the collaborative processes of free open source (mathematical) software development so as to produce guidelines for best practice as well as to develop ideas for extending existing processes to a "system of systems".

Description

"Crowdsourcing"—fine-grained collaborative development of ideas, proofs or software—is a common theme to both objectives. The purpose of a VRE is to allow effective crowdsourcing of computationally supported mathematical results (theorems, proofs, etc.), while free software development is inherently a collaborative process, and we wish to study the best ways of allowing it to scale.

In a sense, mathematics has been a crowdsourced endeavour, dating as far back as the foundation of the Royal Society (UK) in the seventeenth century. The first scientific journals were published collections of letters received, posing questions and observations and offering solutions. Although limited by the speed of physical post, this model had much in common with the public email lists that underpinned collaborative software development in the 1990s.

In recent years, the internet and critical tools such as distributed version control have supported much more widespread and finer-grained forms of crowdsourcing, first in software development, and, more recently, in mathematics: examples are provided by online mathematics communities, such as Math-overflow [25] and Polymath Projects [34, 39]. Supporting and encouraging "Mutual crowdsourcing" is the main driving force for developing and maintaining any large-scale open-source virtual research environment.

In this work package we will build on the work of Prof. Martin and her collaborators on the EPSRC project, and, in particular, their study of crowdsourcing, and integrate their findings with tools provided by the burgeoning field of algorithmic mechanism design in order to to optimise crowdsourcing workflows in open-source VREs.

T7.1 Social Science Input to Design Sites: UO (lead), UO, UPS

The purpose of this task is to ensure that the design of OpenDreamKit VRE reflects the lessons learned by social scientists studying the ways in which mathematicians actually collaborate and work. Since UO and Martin in particular are already central in the community working in this area, we are well placed to ensure that this happens.

As soon as the project begins, team members at UO will combine their own work with a review of the published literature, and identify and meet with key research groups in this area, in order to distill relevant current knowledge for use in the design phases of other parts of the project. They will present the lessons learned at project meetings and workshops and deliver it as a report D7.1 in month 3.

After that, they will monitor the further development of this area and ensure that any new insights are communicated promptly to the rest of the project. This will be synthesised for archival purposes into two further reports D7.4 and D7.7 in months 24 and 42.

We will survey the data needed to assess development models of large-scale academic open-source projects, such as 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, to be published as a report (and possibly a conference publication) D7.2. The latter will require non-trivial amount of programming work, even only for the test system, `Sage`.

T7.2 Implications of VREs for Publication M12-M42 Sites: USHEF (lead), UO

A key aspect of the OpenDreamKit VRE is support for the full life-cycle of mathematical research, up to, and after publication of results. While it is necessary to support established models of publication, which are central to mathematical practice and academic life, it is also appropriate to explore whether VRE technologies may enable novel models for the distribution of scientific output that are more effective for new forms of mathematical results.

The current model for dissemination 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 widely used for distribution of documents reflects the *status quo* that a scientific paper is 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 constraints of the printed format are relaxed; however, these dissemination routes lack the formalization that ensures (usually) fair attribution of ideas and commentary.

We will prototype and evaluate tools and ideas for dissemination of scientific knowledge that do not rely on a static format and allow for the full spectrum of scientific debate. The tools will enable proper credit allocation by encouraging shared attribution of ideas, software and data. This will interact with work in WP6 concerning attribution and citeability for mathematical databases.

Tools to be prototyped include live posters for distribution of knowledge, designed for integration with either large touch screens or smaller tablets D7.6 as well as extensions of the Jupyter project that would provide facilities for commenting on notebooks, which we expect to encourage debate on mathematical and computational ideas D7.5.

T7.3 Mechanism Design for Free Software Development M12-M42 Sites: UO (lead)

While crowdsourced open-source software development has become an incredibly powerful force in recent years, it still has limitations. Open source projects tend to be fragile, in the community sense, and suffer from disagreements that ultimately result in "forks" and the resulting duplication of effort. We will analyse this phenomenon in the framework of algorithmic game theory, and try to design finely tuned systems of incentives and rewards for contribution so as to increase the stability of the community and its useful output.

We will focus on three areas: (1) prioritisation of bug fixes and feature requests to achieve reliable and useful systems; (2) effective cooperation among multiple collaborative projects; and (3) making decisions about the strategic direction of the system.

We will use prioritisation as a testbed for designing incentives that encourage all participants to contribute towards sustained development of the most important parts of the system. To this end, we will use ideas from the burgeoning field of mechanism design [27] and in particular recent research on crowdsourcing in algorithmic mechanism design [35]. While doing so, we will apply outcomes to a case study system—`Sage`.

The reason why prioritisation poses a challenge in the development of open-source academic software is that this process is task-driven: typically, tasks (also known as tickets) are posted on a website, and their priorities are set in an ad hoc manner. This model is usually good enough for simple bug fixing, but for more elaborate tasks it often leads to unacceptable delays. We will apply preference and opinion aggregation techniques [4] to develop a community prioritisation scheme for bugs and feature requests (which may rely on a reputation scores technique, such as one used on MathOverflow), and implement this scheme as a TRAC [40] add-on D7.3. As `Sage` is using the TRAC server [3], this will be easy to test on our testbed system.

Trusting results of computer calculations 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 sometimes fall into the short-sighted trap of hiding bugs from potential and current users [8]. Open source systems are only marginally better in this respect, as indicated by recent computer security scares, such as the one around Bash [6]. A game-theoretic analysis of this situation will be attempted.

A key strength of free and open-source software models is the ability to build upon pre-existing software. `GAP`, `PARI/GP`, `Singular` and especially `Sage` have made heavy use of this ability. Problems arise over time, however, as priorities of the system developers diverge. For instance, bugs reported by so-called "downstream" systems may not be given the same priority as bugs reported by direct users of the "upstream" system, or ignored altogether; similarly, incompatible changes can be made as long as they are acceptable to the direct user community, even if they cause problems for a dependent system. We will explore how sociological and game-theoretic insights can be used to reduce these problems.

The results of this task will be summarised in D7.9 and reported at relevant AI workshops and conferences.

T7.4 Evaluation of Micromagnetic VRE M28-M40 Sites: USO (lead), UO, UPS

We will use the micromagnetic VRE demonstrator (**T4.13**), its dissemination workshops (**T2.8**) 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 (**T2.8**) to gather data.
All results and insights will be summarised in a public document (D7.10) 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: 3, Type: R, Dissem.: PU, Lead: UO) *Report on relevant research in sociology of mathematics and lessons for design of OpenDreamKit VRE, part I*
D7.2 (Due: 18, Type: R, Dissem.: PU, Lead: UO) *The flow of code and patches in open source projects*
D7.3 (Due: 24, Type: OTHER, Dissem.: PU, Lead: UO) *TRAC add-on to manage ticket prioritisation*
D7.4 (Due: 24, Type: R, Dissem.: PU, Lead: UO) *Report on relevant research in sociology of mathematics and lessons for design of OpenDreamKit VRE, part II*
D7.5 (Due: 24, Type: DEM, Dissem.: PU, Lead: USHEF) *Demo: Mechanism for comments on posted Jupyter notebooks*
D7.6 (Due: 36, Type: DEM, Dissem.: PU, Lead: USHEF) *Demo: Jupyter Notebook Live Poster*
D7.7 (Due: 42, Type: R, Dissem.: PU, Lead: UO) *Report on relevant research in sociology of mathematics and lessons for design of OpenDreamKit VRE, part III*
D7.8 (Due: 42, Type: R, Dissem.: PU, Lead: USHEF) *Review of new publication mechanisms, including evaluation of demonstrator projects*
D7.9 (Due: 42, Type: R, Dissem.: PU, Lead: UO) *Game-theoretic analysis of development practices in open-source VREs*
D7.10 (Due: 48, Type: R, Dissem.: PU, Lead: USO) *Micromagnetic VRE environment evaluation report*

3.2 Management Structure and Procedures

3.2.1 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. A 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, who 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.

3.2.2 Organizational structure and decision-making

The organizational structure, shown in the Figure XXXXXXX, has been designed to enable efficient coordination of the OpenDreamKit project — the development and evaluation of a VRE toolkit 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 acts as an intermediary between the Partners and the European Commission. The coordinator will oversee the project planning, monitor that execution is carried out in time and that the objectives are achieved and closely interact with the project officer for project monitoring and delivery of the performance indicators. The Project Manager will ensure efficient day-to-day management of the project, reporting, feedback to partners on administrative, financial and legal issues, tracking of resource allocation and consumption, and 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. Clear distribution of tasks, efficient decision making mechanisms and a sound financial management will safeguard the achievement of the project's objectives.

3.2.3 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: The 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. Preparation of scientific and financial progress reports;
3. Controlling Work Package progress by assessing it through technical reports developed by the 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 the strategic orientation of the project. It takes decisions on scientific directions, re-allocation of resources, consortium changes and 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 SC 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 and external organizations, 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: at the request 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 the kick-off meeting of the project.

Responsibilities: to monitor the quality of the Deliverables, their 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.

3.2.4 Project management tools and procedures

Project partners and management bodies will communicate through a dedicated project web platform, maintained by the Project Manager. WP leaders will monitor progress of participants of their WP at least monthly, 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 of ensuring 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 EUG, working in close cooperation with the Project Coordinator, will ensure efficient innovation management. They will carefully monitor new opportunities in order to give, if necessary, new directions 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 15 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 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.

3.2.5 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 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.

Risk	Level with-/without mitigation	Mitigation measures
Recruitment of highly qualified staff	High/Medium	Great care was taken identifying pool of candidates to hire from, and co-ordinating 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.
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.
Implementing infrastructure that does not match the needs of end users	High/Low	Most of the members of the consortium are themselves end-users with a diverse range of needs and points of views; hence the design of the proposal and the governance of the project is naturally steered by demand; besides, because we provide a toolkit, users have the flexibility to adapt the infrastructure to their needs.
Lack of predictability for tasks that are pursued jointly with the community	Medium/Low	The PI's have a strong experience managing community-developed projects where the execution of tasks depends on the availability of partners. Some tasks may end up requiring more manpower from OpenDreamKit to be completed on time, while others may be entirely taken care of by the community. Reallocating tasks and redefining work plans is common practice; this anyway happens to cater for a fast evolving context. Such random factors will be averaged out over the large number of independent tasks.

Figure 7: Initial Risk Assessment

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. An initial risk assessment appears as figure 7.

3.3 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, Warwick, 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 experience of `SageMathCloud` as a forerunner of the types of systems we want to build, the informal involvement of this US group is adding great value to our project. They are keen 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 (Southampton, 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 in 3.1.6, but in general terms:

- Groups 1, 2, and 3 (from the list above) will collaborate 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.

- 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. Throughout the consortium, groups have substantial experience in open source code development.

- 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.

3.4 Resources to be Committed

3.4.1 Guidelines for travel, dissemination and outreach expenses

The community-building nature of this grant proposal requires a large number of staff exchanges, workshops with project partners, as well as 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 computational science-focused communities and their conferences, as well as the domains benefitting from OpenDreamKit, such as Mathematics and Physical Sciences.

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 of a project partner, for instance for coding sprints and one-to-one research visits. We expect a similar cost per week while hosting visitors. For the half-yearly 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 the researcher in total to do 4 one-week visits to other sites (each at €750) every year, totaling €12000 over 4 years).

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.

Outreach costs We also request €1000 per year per partner (several partners have other means do pay these costs, and for them these are not needed) 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 workshop 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 and 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 this recent workshop http://wiki.sagemath.org/days57.

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

.

WP	Title	UPS	Logilab	UVSQ	UJF	CNRS	UO	USHEF	USO	USTAN	UW	JacU	UK	US	UZH	Simula	total
WP1	Management	*28*	2	2	2	2	2	2	2	2	2	2	2	2	1	2	**55**
WP2	Community Building/Dissemination	*10*	6	2		20		17	16	18			2	30		2	**123**
WP3	Component Architecture	64	14	*8*	6		4		6	16							**118**
WP4	User Interfaces	26	12	2		28		6	16	18		12	2	4		*28*	**154**
WP5	High Performance Math. Computing	6	12		*52*	40		12		18			60				**200**
WP6	Data/Knowledge/Software-Bases	4	2							10	25	*46*			12		**99**
WP7	Social Aspects						*23*	18	6								**47**
	totals	**138**	**48**	**14**	**60**	**90**	**29**	**55**	**46**	**82**	**27**	**60**	**66**	**36**	**13**	**32**	**796**

Efforts in PM; WP lead efforts light gray italicised

Figure 8: Summary of Staff Efforts

3.4.2 Resource summaries for consortium member sites

In this section we briefly describe the requested resources. See the participant descriptions in the description of the consortium for the specific role of each member.

Resources Université Paris Sud Paris Sud requests 12 person months for the project coordinator (Nicolas M. Thiéry), 6 person months for two researchers (Florent Hivert, and Samuel Lelièvre) and for the lead PI (Viviane Pons), 48+36 months for two full time developers, and 24 months for a part time project manager for the full duration of the project.

PS	Cost (€)	Justification
Travel	103,000	Travel (see 3.4.1)
Publication charges	1,000	Open access publication charges (see 3.4.1)
Other goods and services	82,000	8 developer workshops **T2.3**, and 4 training and dissemination workshops **T2.1**
Total	186,000	

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

Resources Logilab Logilab requests 36 person months for its engineers (Julien Cristau, Florent Cayré and Olivier Cayrol). They will bring their expertise (database design, software architecture, computer-domain knowledge) to numerous tasks, and will notably contribute to the packaging of `Sage` (**T3.3**), the enhancement of existing forges (**T3.7**) and the addition of HTML5 widgets in notebooks (**T4.5**).

Logilab requests 12 person months for subcontracting an engineer (Serge Guelton), main developer of `Pythran`, that will work on **T5.7**.

LL	Cost (€)	Justification
Travel	17,200	Travel (see 3.4.1)
Total	17,200	

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

Resources Université de Versailles Saint-Quentin Université de Versailles Saint-Quentin requests 12 person months for the lead PI (Luca De Feo) and 2 person months for a researcher (Nicolas Gama). Because of its small size and geographical proximity to Paris Sud, Université de Versailles is not going to hire any full-time personnel for the project.

UV	Cost (€)	Justification
Travel	12,600	Travel (4 EU conferences, 4 one week visits to project partners, 12 project meetings)
Total	12,600	

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

Resources Université Joseph Fourier UJF requests 12 person months for an engineer (Pierrick Brunet) starting in month 0 and working on task **T5.7**, 24 person months for another engineer starting in month 12 and working on task **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.

UJF	Cost (€)	Justification
Travel	60,850	Travel (see 3.4.1)
Publication charges	4,000	Open access publication charges (see 3.4.1)
Equipment	24,000	A large multicore server with multiple accelerators to experiment heterogeneous computing; 2 laptops
Other goods and services	30,000	2 developer workshops: HPC and Pythran
Audit	4,000	Audit cost
Total	122,850	

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

Resources CNRS Aquitaine CNRS Aquitaine requests 12 person months for the lead PI Vincent Delecroix and the PARI/GP head Karim Belabas, 6 person months for the other PI (Bill Allombert, Adrien Boussicault and Loic Gouarin) and a 48 person months for a full time developer working on tasks **T5.1**, **T5.6**, **T4.10** and **T4.12**.

UB	Cost (€)	Justification
Travel	56,700	Travel (see 3.4.1)
Publication charges	1,000	Open access publication charges (see 3.4.1)
Other goods and services	111,000	1 HPC workshop **T2.3**, 4 Ateliers Pari **T2.3**, 4 dissemination workshop in developing countries **T2.5**
Audit	4,000	Audit cost
Total	172,700	

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

Resources University of Oxford University of Oxford requests 24 person months for a co investigator (Dmitrii Pasechnik), who will be employed for the whole duration of the project, 2.4 person months for the lead PI (Ursula Matrin) and 2.4 person months for the other co investigator (Edith Elkind) to work on WP7, of which 2 person months for WP1 (Management). From his total involvement, Pasechnik will take 4 person months to work on WP3. Martin and Elkind both hold personal fellowships (funded by EPSRC (UK) and by ERC, respectively) on topics closely related to the project, enabling them to take part in the project only at a fraction of the full cost; however, they will need funding to travel to project meetings. Open access publication charges will be met by the host institution.

UO	Cost (€)	Justification
Travel	25,000	Travel (see 3.4.1)
Equipment	3,000	Laptop and a large display for a co investigator
Total	28,000	

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

Resources University of Sheffield Sheffield requests 42 person months for a researcher, 6 person months for the lead PI (Neil Lawrence) and 6 person months for the co-Investigator (Mike Croucher). The lead PI will take on all management responsibilities. Two researchers will be employed, one for a shorter period of 6 months, with a specific focus on the SGE Implementation **T5.8**. The other for a period of 36 months, with appointment at month 6, whose focus will be on user interfaces, social aspects **T7.2** and course development **T2.6**. 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. One person month of an administrator is requested for assistance with workshop and meeting organisation etc.

Sheffield will host one of the project main workshops (15,000) and will also host two workshops on machine learning and data science with OpenDreamKit. We also request equipment for workshop recording at 2,000, and funding for a large touch screen for live notebook posters (D7.6) at 5,000. For the researchers we request screens and computers (3,000).

USH	Cost (€)	Justification
Travel	53,200	Travel (see 3.4.1)
Publication charges	4,000	Open access publication charges (see 3.4.1)
Equipment	10,000	High performance laptops and multi touch large screen for **T7.2**
Other goods and services	20,000	Workshops (Project meeting and two dissemination workshops, see **T2.6**)
Other goods and services	5,000	HPC Compute Time (see **T5.8**)
Audit	3,000	Audit cost
Total	95,200	

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

Resources Southampton Southampton requests 38 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.

USO	Cost (€)	Justification
Travel	51,500	Travel (see 3.4.1)
Publication charges	4,000	Open access publication charges (see 3.4.1)
Equipment	10,000	HPC Workstation (6k) to host micromagnetic VRE web server, **T4.14**, and two high performance laptops (2x2k)
Other goods and services	19,600	4 Dissemination workshops overseas (travel for teachers & room hire), **T2.8**
Audit	5,200	Audit cost
Total	86,300	

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

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.

SA	Cost (€)	Justification
Travel	76,800	Travel (see 3.4.1)
Publication charges	4,000	Open access publication charges (see 3.4.1)
Equipment	15,000	Compute servers for parallel software development and testing (tasks **T5.2**, **T3.5**)
Other goods and services	17,200	2 dissemination workshops (room hire and subsistence for external participants; task **T2.3**)
Audit	4,300	Audit cost
Total	117,300	

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

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 WP1 (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.

UW	Cost (€)	Justification
Travel	9,600	Project meetings and partner site visits; investigator co-funded by LMFDB grant (see 3.4.1)
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 13: Overview: Non-staff resources to be committed at UW (all in €)

Resources Jacobs University Bremen JacU requests 6 PM each for the PIs (Prof. Michael Kohlhase leads WP6) and Dr. habil Florian Rabe (theoretical foundations of triform theories). Furthermore, we request 24 PM each for a research programmer (Dr. Christian Maeder) and a junior researchers (Mihnea Iancu M.Sc.). The first will do much of the actually system development in WP4 and WP6 while the latter will concentrate on the cases studies in WP6.

JU	Cost (€)	Justification
Travel	53,600	Travel (see 3.4.1)
Publication charges	4,000	Open access publication charges (see 3.4.1)
Equipment	14,000	for two web/compute servers for **T4.7**; they need 256 GB RAM each for the math search engine from **T6.9**
Audit	4,300	Audit cost
Total	75,900	

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

Resources University of Kaiserslautern UK requests 48 person months for a researcher to work on tasks **T5.4**, **T4.1**, 12 person months for a researcher starting in month 0 to work on task **T5.5**, and 6 person months for the lead PI (Wolfram Decker). The lead PI will take on all management responsibilities.

UK	Cost (€)	Justification
Travel	67,100	Travel (see 3.4.1)
Other goods and services	65,000	5 developer workshops **T2.3**
Audit	2,600	Audit cost
Total	134,700	

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

Resources University of Silesia University of Silesia will include three people in the project and their involvement will 12 person months each. Jerzy Luczka and Jan Aksamit will work on interactive books **T2.9** which will demonstrate the real case of using both Structured Text and interactive features of the VRE. Marcin Kostur will lead and contribute to this task as well.

Marcin Kostur will lead the part of **T4.9** which will be connected with 3d visualization of data produced by the Lattice Boltzmann software 'sailfish'. After initial research work done together with Simula (task leader) will require to subcontract the development of software visualization.

The justification for subcontracting is as follows. We have much experience as users of 3d visualization software for fluid dynamics. However the expertise in Computer Graphics (e.g. WebGL) is not enough at the Department of Mathematics, Physics and Chemistry. Instead of building the expertise it is financially more efficient to specify and outsource the programming task to professionals.

US	Cost (€)	Justification
Travel	18187,50	Travel (see 3.4.1)
Publication charges	0,000	Open access publication charges (see 3.4.1)
Other goods and services	50,000	Subcontracting costs
Audit	3,000	Audit cost
Total	71187,50	

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

Resources University of Zürich Zurich will employ one person associated with the project, Paul-Olivier Dehaye. Twelve person-months will be dedicated to WP6 and spread over the four years, with an extra one for the management (WP1). He will devote additional time to these efforts, paid from other sources (University of Zurich and Swiss Science Foundation).

He will lead tasks **T6.1**, and assist for **T6.3**, **T6.2**, **T6.4**, **T6.6** and **T6.7**. He is in charge of deliverables D6.8, and D6.6.

ZH	Cost (€)	Justification
Travel	26,800	Travel (see 3.4.1)
Publication charges	4,000	Open access publication charges (see 3.4.1)
Equipment	2,000	laptop for investigator
Total	32,800	

Table 17: Overview: Non-staff resources to be committed at the University of Zurich (all in €)

Simula Research Laboratory Simula requests 28 person months for research activities to lead Work package 4 and contribute to its specific tasks. We also dedicate 4 person months for management as well as dissemination and communication activities (to be split equally). These activities will be mainly performed by the lead PI with the extensive support of the local management team.

SR	Cost (€)	Justification
Travel	56,200	Travel (see 3.4.1)
Publication charges	4,000	Open access publication charges (see 3.4.1)
Other goods and services	8,000	Organisation of the Jupyter workshop
Audit	3,500	Audit cost
Total	68,200	

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

3.4.3 Resources Summary

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

Other direct cost items

OpenDreamKit	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 national and international (European Inventor Award 2013, Wolf Prize 2010, Holweck Prize 2009, Japan prize 2007) prizes. The Université Paris-Sud offers a complete range of qualifications, 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 commercialisation 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 (**T3.6**, **T3.2**, **T4.1**, **T4.4**, **T4.5**, **T5.6**, **T6.8**). 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 starting research career, she worked for two years 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 in `Sage`. She is heavily involved in the promotion of `Sage`, participating in `Sage` Days and running `Sage` introduction tutorials or `Sage` presentations at 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 `Sage` tickets (co)authored and as many refereed, Hivert is himself a core `Sage` developer, with contributions including key components of the `Sage` infrastructure (documentation, automated test, combinatorics infrastructure, parallelism, ...) and specialised 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 co-organised 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 organised community of Sage developers, gaining a leading position in its field, and making a major impact on one hundred publications[8]. Along the way, he coauthored part of the proposal for NSF `Sage-Combinat` grant OCI-1147247, and co-organised or taught at a dozen training and dissemination actions (workshops, summer schools, etc.), in America, Africa, Europe and 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), specialised research libraries (e.g. root systems), thematic tutorials, and two chapters of the book "Calcul Mathématique avec `Sage`".

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.

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

He uses and actively promotes Sage since 2010. He is in the top 15 contributors of the AskSage questions and answers forum. He co-organised 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).

NN (res, 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 during the whole duration of OpenDreamKit.

Publications, achievements

1. Leadership in 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. Contributing of more than 500 tickets to Sage.

Previous projects or activities

1. Hosting six Sage Days (week-long training and development workshops).
2. Co-organising six other Sage Days.
3. Founding and regular organisation of bimonthly Sage User Group meetings 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 are 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 visualisation 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 employs over 20 engineers and PhDs proficient in software engineering, knowledge representation, design and management of IT infrastructure, and other areas.

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 visualise complex mathematical properties, fluid workflow tools to ease large-scale collaboration, etc.

Curriculum vitae

Florent Cayré (PI, male, 6 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 designing the numeric tools for the turbines design and then was head of the group “Méthodes de conception de turbines ; aérothermique et combustion” performing collaborative R&D programs and development and integration of various numeric tools (`Python`, `C++`, `C`, `Fortran`). He co-funded the SecondWeb company tat carries out the development of complex Web applications based on `CubicWeb` platform.

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

Olivier Cayrol (lead PI, male, 6 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 software products such as `Sphinx` or `reportlab`, and defines numerous extensions for answering the specific needs.

David Douard (PI, male, 6 PM) holds a Master Degree from the Ecole Nationale Supérieure de Physique de Strasbourg and a PhD in computer science from Université Paris VI.

As a PhD studend, he developed a graphical user interface for the computational system he used for his research as well as lots of processing and visualization tools.

He worked two years at EDF where he was the lead developer of the software in charge of evaluating financial risks on the energy market with a specific, intensive work on the graphical user interface and on the code driving the simulations.

He has been working at Logilab since 2006, building complex scientific applications involving the management and visualization of large amounts of data. He has trained tens of engineers and researchers to `C/C++`, `Python`, GUI libraries (`Tk`, `PyQt`, `wxPython`), `Fortran` and scientific computing.

Julien Cristau (PI, male, 18 PM) holds Master Degree and PhD from University Paris VII, where he carried out mathematical research in automata and linear games. As a `Debian` developer since 2007, he maintained its key components such as the `X11` windowing system and acted as a `Debian` Release Manager since 2011.

Working as a software engineer in the R&D department of Logilab since 2011, he developed software using many different languages and systems, helped to release and distribute software on many different platforms, maintained parts of the infrastructure and trained other people.

Serge Guelton (R, male, 12 PM) Serge Guelton holds an engineering degree in Computer Science and telecommunication from from Télécom Bretagne and a PhD in compilation and parallelism. He's been working as an expert engineering in various INRIA teams, and as a lead developer in several start-ups.

He's the lead developer of the `Pythran` project, a `Python`-to-`C++` compiler for high-performance scientific kernels.

Publications, products, achievements

1. CubicWeb is a semantic web framework that is available under the LGPL license and received a DataConnexion prize from Etalab (the french government team dedicated to Open Data)
2. Logilab has been contributing to free software since its creation in 2000 and is known for it in France and several other countries. It authored `PyLint`, the static `Python` code checker used worldwide, and has always had at least one `Debian` Developer on staff, thus supporting the largest free software distribution used by millions of people.
3. At OBHM 2013, the 19th Annual Meeting of the Organisation for Human Brain Mapping, Logilab presented a poster which explains the work done using CubicWeb on brain imaging and genetics data in collaboration with INRIA, INSERM and the CEA during the Brainomics project co-financed by Agence nationale de la Recherche.

Previous projects or activities

1. Logilab was a member of the consortium for the ASWAD EU project, with a role of software developer. The project demonstrated the interest of free software for setting up workflows in public administrations.
2. Logilab was a member of the consortium for the KIDDANET EU project, with a role of software developer. The project implemented a proxy to protect kids browsing the internet.
3. Logilab was a member of the consortium for the PYPY EU project, with a role of software developer. The project implemented a `Python` interpreter in `Python`, to explore new ways to compile and optimise the execution of `Python` code.
4. Logilab was a member of the consortium for the OpenHPC french project, with a role of software developer. The project advanced the state of the art of free software for high performance simulation.
5. Logilab was a member of the consortium for the BRAINOMICS french ANR project, with a role of software developer. The project advanced the state of the art of shared databases for the brain imaging and genetics data.
6. Logilab was a member of the consortium for the THALER french ANR project, with a role of software developer. The project advanced the state of the art of free software for high performance simulation of molecular dynamics.

Significant infrastructure

Logilab is maintaining its own infrastructure, using virtualisation techniques and tools such as `OpenStack`, `SaltStack`, `Docker` and others.

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, Modelling 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, Modelling 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/GP`, `Maple`, `Magma`, and others). 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, 12 PM) 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 organises various events for the introduction of `Sage` to beginners and young researchers.

Nicolas Gama (R, male, 2 PM) 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 specialised in Algebraic computations, cryptology, codes and hybrid symbolic-numeric dynamical systems. The MOAIS team is specialised 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 system `XKAAPI`. Relying on this expertise, the UJF site will be coordinating the tasks related to High Performance Mathematical Computing, forming Work Package 5: **T5.1**, **T5.2**, **T5.3**, **T5.4**, **T5.5**, **T5.6**, **T5.7**, **T5.8**. It will also be in charge of the task **T5.3** on the development of high performance exact linear algebra with the `LinBox` library.

Curriculum vitae

Clément Pernet (leadPI, male, 15 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-organised 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, 9 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-organised 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` runtime libraries.

With about 25% of all commits in the `Pythran` [4.1.4] project, Pierrick is one of the core developers of this project which compiles a subset of the `Python` language to native `Python` modules.

Unknown (res, 24 PM) Full time developer on `LinBox`, and its integration into `Sage` in the framework of Task **T5.3**. 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 systems: `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 multi-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. Organisation of the ISSAC'12 conference, the main international conference in computer algebra, and of PASCO'15, the ISSAC'15 satellite conference on parallel computer algebra.

4.1.5 CNRS: CNRS Aquitaine (FR)

Bordeaux is an important center of studies and research in France with approximately 50,000 students, 2,000 PhD students and 5,000 researchers. The University of Bordeaux was founded in the 15th century and nowadays the city regroups two universities, dozens of schools and 100 research laboratories with partner institution such as CNRS, Inserm, INRA and INRIA.

The Institut Mathémathiques de Bordeaux (IMB) is a leading institution in Number Theory. It is the home of the software PARI/GP and the Journal de Théorie des Nombres de Bordeaux.

The city of Bordeaux also hosts two important young laboratories for computer science: Laboratoire Bordelais d'Informatique (LaBRI) and INRIA-Bordeaux.

In the context of this proposal, Bordeaux has a long standing experience in Algorithmic Number Theory and two significant hardware infrastructures (Plafrim and Avakas). The CNRS Aquitaine main task in this project is the extension of PARI/GP in relation with the other software components and the .

The CNRS Aquitaine site will lead the development of PARI/GP within workpackages WP4,WP5, WP2. It will also play an important role in WP5 with respect to combinatorics and the organisation of workshop in favour of developing countries in WP2.

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 to Sage with 30 tickets authored and around 50 reviewed. He organised several Sage days and Sage workshops in Bordeaux, Marseille, Orsay, Perpignan, Bobo Dioulasso (Burkina Faso), Saint-Louis (Sénégal).

Karim Belabas (PI, male, 12 PM) is a Professor of Mathematics in Bordeaux (France) since 2005. He 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 Medalist), and edited the book "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 € 200k (ALGOL project, 2007–2011) and was part of a €2.5m 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-)organised 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) is a research engineer in Bordeaux. He has a great expertise in algorithmic number theory and is one of the main developer of PARI/GP. He is the developer of the GP2C compiler to convert GP code into pari code, and the GALPOL database (database of polynomial with prescribed Galois groups).

He also has 5 articles in international journals and is a regular contributor of the software GAP and the Debian project.

Adrien Boussicault (R, male, 6 PM) Maître de Conférences at the LaBRI (Laboratoire Bordealais de Recherche en informatique), Adrien Boussicault is a young researcher in Algebraic and Enumerative Combinatorics. He has 8 papers in international journals. His contributions to Sage include writing 3 tickets to implement combinatorial objects and co-organising Sage-Combinat Days 57.

Loïc Gouarin (R, male, 6 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 organising conferences, meetings, and seminars. In this context, he organises himself 3 to 4 training and development workshops per year, and promotes the use of `Python` for teaching and research in France.

Organisationally, due to purely administrative constraints within CNRS, Loïc Gouarin will be attached to the CNRS Aquitaine.

NN (R, 48 PM) We will hire a research engineer to work at Bordeaux under the leadership of Prof. Karim Belabas and Dr. Vincent Delecroix on the tasks of WP5, WP3 and WP4. He or she will work on the following tasks:

- parallelisation of some low-level algorithms in `PARI/GP`,
- creation of `Cython/Python` bindings for `PARI/GP`,
- implementation of a mixed `C/Python` library for iteration of combinatorial objects and its integration in `Sage`,
- implementation of the functionality to output combinatorial objects in `Sage` with the support of various possible options (raw text, pretty-printing, export to LaTeX, `tikz`, `matplotlib`, etc.).

Publications, products, achievements

Some recent publications :

1. Karim Belabas, Eduardo Friedman, *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. Karim Belabas et al. *Explicit methods in number theory. Rational points and Diophantine equations*, 179 pages, Panoramas et Synthèses 36, 179p., 2012.
4. Bill Allombert, Yuri Bilu and Amalia Pizarro-Madariaga, *CM-Points on Straight Lines*, to appear in "Analytic Number Theory" (dedicated do H. Maier), Springer.
5. Vincent Delecroix, *Cardinality of Rauzy classes* Ann. Inst. Fourier, 63 no 5 (2013), p. 1651-1715.
6. Jean-Christophe Aval, Adrien Boussicault, Mathilde Bouvel, Matteo Silimbani *Combinatorics of non-ambiguous trees*, Advances in Applied Mathematics 56 (2014), p. 78-108.

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

1. The Plafrim is a regional federation hosted at INRIA Bordeaux (in partnership with the LaBRI and IMB). It has an important cluster devoted to experimental code (1188 cores).
2. The Mésocenter de Calcul Intensif Aquitain (MCIA) is localised in Bordeaux. It hosts the Avakas cluster (3328 cores, 38 TFlops) and the M3PEC cluster (432 cores).

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 `Sage`.

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 characterised 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 crowd-sourced 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 Associate 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-organised 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 optimisation, 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=2484920.2485003
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 `Sage`, 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 Research 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 assessment 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 by 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.

The University of Sheffield will focus on tasks on HPC (**T5.8**), dissemination (**T2.6**) and social aspects of the project (**T7.2**). It will host one of the main project meetings and run regular workshops on data science and Gaussian processes which incorporate OpenDreamKit outputs.

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 personalised 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 was 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 was 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 implementations available, most recently using `Python` and `IPython` as the main medium for communicating their work. Their Gaussian process python software framework[9] 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.[10]

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.

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.

NN (R, 36 PM) We will hire a post-doctoral research fellow to carry out the work at Sheffield in collaboration with Neil Lawrence and Mike Croucher. The fellow will ideally have a background in data science, 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 skills in the development of collaborative educational materials.

[9] https://github.com/SheffieldML/GPy

[10] http://ml.dcs.shef.ac.uk/gpss/

NN (R, 6 PM) We will fund existing post-doctoral research resource to carry out work on implementing `Jupyter` on SGE for ease of HPC performance. They will work collaboratively with Neil Lawrence, Mike Croucher and the other Sheffield researcher appointment to achieve this goal. The fellow will have a background in computational science, with experience of HPC and `Jupyter` Notebook experience. As is normal we will require good communication and team working skills.

NN (res, 1 PM) We will support one month of an administrators time for assistance with workshop organisation and project administration across the whole duration of OpenDreamKit.

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 was 3-4 days long.

Significant infrastructure

1. The Sheffield Institute for Translational Neuroscience is a £18m world leading institute for research into neurodegenerative disorders. It houses clinicians, biologists and computational scientists 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.
2. The Sheffield group are regular contributors to open source software including a python framework for Gaussian process modelling (https://github.com/SheffieldML/GPy), contributions to the Bioconductor framework for computational biology (puma, tigre, gprege), and more recently frameworks for open data science (https://github.com/sods/ods).
3. The Sheffield group has taken a leading role in education in data science and machine learning with a series of workshops and summer schools which use `Jupyter` as the main interface for practical implementation of ideas (http://ml.dcs.shef.ac.uk/gpss/).

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 € 1m 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 centre 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 provides a finite-element micromagnetic simulation suite to a community of material scientists, engineers and physicists who research magnetic nanostructures in academia and industry. He has designed the package in 2005 so that it has an `IPython`-compatible `Python` interface, to make the workflow of using the simulation package as accessible as possible to scientists without substantial computational background. He has extensive experience in micromagnetic simulation tool development and use, and due to this an outstanding understanding of the requirements for computational workflows in this micromagnetic research community.

He has deep interest in excellence and innovation in learning and teaching. He has been awarded the prestigious Vice Chancellor's teaching award ($\pounds 1000$) 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, 38 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 educational materials to the best support of 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 "Iridis 4", the largest university owned Supercomputer 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 University of St Andrews is the third-oldest in the English-speaking world (founded 1413) and pursues cutting edge research in all its academic Schools and interdisciplinary Institutes and Centres supported by a wide range of grants from many sources. The University currently has about 8000 students and 600 academic staff in 19 Schools.

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. CIRCA includes 36 staff and 20 research students.

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[11]. 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. He is active both 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.
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/).

[11] Applicable Algebra and Error Correcting Codes

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 (see 1.3.9).
3. **HPCGAP: High Performance Computational Algebra and Discrete Mathematics** EP/G055181 2009–2014, £1.5m, four sites, coordinator (see 1.3.9).

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 [17]) Group headed by *Prof. Dr. Michael Kohlhase* specialises 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) 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 more than 10 contributors and currently consists of more than $30,000$ lines of `Scala` code.

He served in the organising committee of 2 and the program committee of 6 international conferences (2 as track chair) on intelligent computer mathematics, and has organised 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 software developer at Jacobs University. He has extensive experience in designing and implementing logic-based software systems. 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` system.

Relevant previous experience:

The KWARC group is the lead implementor of the OMDoc (Open Mathematical Document) format for representing mathematical knowledge [18] and redeveloped its formal core in the OMDoc/MMT format [33]. The latter has been implemented in the MMT system [32, 33] 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[22, 21].

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 [15] as well as automatic annotation using linguistic techniques [12]. 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 [23, 24, 13].

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:

- Modelling 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.

JacU lead WP6: Data/Knowledge/Software-Bases and tasks related to active and structured documents as user interfaces (WP4: User Interfaces) to the math VRE to be developed in OpenDreamKit. It will run an infrastructure for authoring and interacting with such documents and a search engine for the $\mathcal{DKS}$ base.

4.1.12 UK: University of Kaiserslautern (DE)

The University of Kaiserslautern (UK) is a medium sized university founded in 1970. It currently consists of 12 departments, ranging from mathematics and business studies and economics, computer sciences and electrical and computer technology over mechanical and process engineering, biology, chemistry and physics to architecture, regional and environmental planning, and social sciences. The university has 13,725 students, 3636 of whom are remote study students. The scientific location of Kaiserslautern is also distinguished by the presence of multiple external research institutes of considerable reputation, including two facilities of the Fraunhofer network and the German Research Institution for Artificial Intelligence. All these institutions maintain close links and even share staff with the corresponding departments of UK, which is chairing the Science Alliance Kaiserslautern, a network of these research institutions. The university conducts a number of international collaborations and successfully participated in projects funded under several EU Framework Programs and has gathered comprehensive experience both as coordinator and partner in research networks and projects. Besides projects with national funding, the UK is also very active in the field of international research. In this context, the funding instruments available in the EU Framework Programmes play an important role. In total, the UK is partner to a total of 11 projects (as of January 2015) conducted under the 7th FP and Horizon 2020. Nine further individual projects funded by ERC (2) or Marie-Curie measures (7) are being co-ordinated by researchers at UK. By this involvement to date, UK has procured more than 13 million Euros under the 7th FP.

In the context of this proposal, the *Algebra, Geometry, and Computer Algebra Group* of the Department of Mathematics at UK is widely known for its long tradition in computational algebraic geometry and algebra, with particular emphasis on the development of the computer algebra system `Singular` and its satellites and subsystems such as `Factory`, `PolyBori`, and `Plural`. Kaiserslautern's main tasks in this project are to add very fine-grained parallelism to some key components of `Singular` and to work on the maintenance and improvement of `MPIR`.

Curriculum vitae

Prof. Dr. Wolfram Decker (leadPI, male, 6 PM) 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 30 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-)organised 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 cross-linking `Singular` to other systems, most notably to `GAP`, and parallelising `Singular`. These tasks are fundamental to the OpenDreamKit project.

Dr. William Hart (PI, 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 of the famous Birch and Swinnerton-Dyer conjecture, one of the Millennium Prize Problems listed by the Clay Mathematics Institute) for 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 `Macaulay2` 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.

NN (res, 48 PM) We will hire a full time experienced software developer to work under the leadership of Wolfram Decker on adding very fine-grained parallelism to some key components of `Singular`. The fellow will have past experience of parallelism in software development. We further require good communication and team working skills.

NN (res, 12 PM) We will hire a full time highly specialised software developer and assembly expert, to work under the leadership of William Hart on the performance task **T5.5** for `MPIR`.

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-Centre of the TU Kaiserslautern.

4.1.13 US: University of Silesia (PL)

The University of Silesia in Katowice was established in 1968. Now, with 12 faculties and several interdisciplinary schools and centres, 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. He is the 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 computational 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: €263k, 2014-2017)
- “Computers in Science Education: iCSE” http://icse.us.edu.pl (budget: €1m, funded by EFS, 2011-2014)

He is also co-author and a task coordinator of PAAD (Platforma Analiz i Archiwizacji Danych) project funded by POIG program for 2014-2015 with a total budget of €4m. 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 (http://zft.us.edu.pl/luczka) is a full professor of physics at the University of Silesia (Katowice, Poland) and the Head of the 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 an 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-organised 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 the University of Silesia as a research assistant, lecturer and then senior lecturer. His skills combine 40 years 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 to use his vast lecturing experience for the creation of interactive demonstrators of OpenDreamKit.

Publications, products, achievements

1. Sailfish: A flexible multi-GPU implementation of the lattice Boltzmann method, Computer Physics Communications Vol.181(9), 2350-2368;2014, 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, https://github.com/marcinofulus/CUDASDE.git.
3. iCSE course materials, http://visual.icse.us.edu.pl/iCSE_main/.

Previous projects or activities

Internationalisation 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 involved in the implementation of several FP6 and FP7 projects:

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) - € 1m grant from European Social Fund, 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) € 3.8m, funded is mostly HPC centre for research with interactive access based on web based notebook UI.
3. 2014-30.11.2015 CNS: Centre of Applied Science - 2nd stage, Infrastructure grant includes € 0.5m funding for small HPC and cloud infrastructure for education. Technically this will be extension of research HPC centre for educational purposes.

Significant infrastructure

The University of Silesia has finished or currently implements ESF grants totalling to about € 120m for infrastructure, laboratories and computing centres. New HPC centres are under construction (PAAD and CNS projects) which will provide necessary hardware for development and implementation of virtual research environments.

4.1.14 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, 13 PM) 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 specialised 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, which is, at the time of the submission, the sixth largest (and most energy-efficient) supercomputer in the world (and currently being upgraded).

4.1.15 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 emphasises 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 organised 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, Centre 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 centre-oriented schemes are the most prestigious funding instruments offered by the Research Council of Norway.

Curriculum vitae

Hans Petter Langtangen (leadPI, male, 8 PM) Hans Petter Langtangen is director of the Centre for Biomedical Computing at Simula Research Laboratory, a Norwegian Centre 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 modelling, 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.

His scientific output 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.

NN (R, 24 PM) We will hire a post-doctoral senior research fellow to carry out the work at Simula, under the leadership of and together with Hans Petter Langtangen. The fellow will have a background in computational science, combined with profound `IPython` and `Jupyter` Notebook experience, and past experience of software engineering. Ideal candidate will also have good communication skills and team working abilities, and in particular interest and skill in the development of education materials to best support this part of the project.

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..

4. M. S. Alnæs, A. Logg, K. B. Ølgaard, M. E. Rognes, G. N. Wells. Unified Form Language: A domain-specific language for weak formulations of partial differential equations, ACM Transactions on Mathematical Software, 40(2) (2014).

Previous projects or activities

1. The Centre for Biomedical Computing, a Norwegian Centre of Excellence, awarded by the Research Council of Norway (€ 10m, 2007–2017).
2. The FEniCS Project (www.fenicsproject.org, ongoing since 2007).

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.

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

The University of Bordeaux will be involved as a third party. Geographically this is the same entity as CNRS Aquitaine, but it is needed to include staff from the university (i.e. Karim Belabas and Adrien Boussicault).

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. 15, 43).

[2] http://math.nist.gov/oommf/oommf_cites.html (cit. on p. 15).

[3] (Cit. on p. 57).

[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. 57).

[5] T. Sage-Combinat community. *Sage-Combinat: enhancing Sage as a toolbox for computer exploration in algebraic combinatorics*. `http://combinat.sagemath.org`. 2008 (cit. on p. 50).

[6] *CVE-2014-6271*. (cit. on p. 57).

[7] *Docker – An open platform for distributed applications for developers and sysadmin*. hrefhttp://docker.comhttp://docker.com. 2015 (cit. on pp. 42, 47).

[8] A. J. Durán, M. Pérez, and J. L. Varona. "The Misfortunes of a Trio of Mathematicians Using Computer Algebra Systems. Can We Trust in Them?" In: *Notices of AMS* 61.10 (2014). , pp. 1249–1252 (cit. on p. 57).

[9] W. M. Farmer. "Biform Theories in Chiron". In: *Towards Mechanized Mathematical Assistants. MKM/Calculemus*. Ed. by M. Kauers, M. Kerber, R. Miner, and W. Windsteiger. LNAI 4573. Springer Verlag, 2007, pp. 66–79. ISBN: 978-3-540-73083-5. DOI: `http://dx.doi.org/10.1007/978-3-540-73086-6_6`. URL: `http://dx.doi.org/10.1007/978-3-540-73086-6_6` (cit. on p. 53).

[10] 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).

[11] J. Fromentin and F. Hivert. "Exploring the tree of numerical semigroups". In: (2013). Hal: `00823339v2` (cit. on p. 51).

[12] 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. 82).

[13] 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. 82).

[14] *Interactive Python (IPython)*. http://ipython.org. 2014 (cit. on p. 11).

[15] 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. 82).

[16] K. Ketcheson (cit. on p. 11).

[17] *Knowledge Adaptation and Reasoning for Content*. URL: `http://kwarc.info` (visited on 05/12/2011) (cit. on p. 81).

[18] 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 pp. 53, 82).

[19] M. Kohlhase and M. Iancu. "Searching the Space of Mathematical Knowledge". In: *DML and MIR 2012*. Ed. by P. Sojka and M. Kohlhase. in press. Masaryk University, Brno, 2012. ISBN: 978-80-210-5542-1. URL: `http://kwarc.info/kohlhase/papers/mir12.pdf` (cit. on p. 55).

[20] M. Kohlhase, F. Mance, and F. Rabe. "A Universal Machine for Biform Theory Graphs". In: *Intelligent Computer Mathematics*. Conferences on Intelligent Computer Mathematics. (Bath, UK, July 8–12, 2013). Ed. by J. Carette, D. Aspinall, C. Lange, P. Sojka, and W. Windsteiger. Lecture Notes in Computer Science 7961. Springer, 2013. ISBN: 978-3-642-39319-8. DOI: `10.1007/978-3-642-39320-4`. URL: `http://kwarc.info/kohlhase/papers/cicm13-uom.pdf` (cit. on p. 53).

[21] 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 pp. 54, 82).

[22] 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. 82).

[23] 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. 82).

[24] *MathHub.info: Active Mathematics*. URL: http://mathhub.info (visited on 01/28/2014) (cit. on p. 82).

[25] *MathOverflow*. URL: http://mathoverflow.net (visited on 09/08/2011) (cit. on p. 56).

[26] *MathWebSearch – Searching Mathematics on the Web*. URL: http://search.mathweb.org (visited on 04/26/2013) (cit. on p. 54).

[27] N. Nisan. “Introduction to mechanism design (for computer scientists)”. In: *Algorithmic game theory*. Cambridge Univ. Press, Cambridge, 2007, pp. 209–241 (cit. on p. 57).

[28] *Nmag — Guided tour*. http://nmag.soton.ac.uk/nmag/0.2/manual/html/guided_tour.html (cit. on p. 47).

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

[30] *Notebook based interactive GUI for common simulation requirements*. https://www.youtube.com/watch?v=RhMHgQbP__A. 2014 (cit. on p. 47).

[31] *Project Jupyter*. http://jupyter.org. 2014 (cit. on p. 11).

[32] F. Rabe. *The MMT Language and System*. URL: https://svn.kwarc.info/repos/MMT (visited on 10/11/2011) (cit. on p. 82).

[33] 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 pp. 53, 82).

[34] 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. 56).

[35] 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. 57).

[36] Software.intel.com. *Intel® Cilk™ Homepage*. https://www.cilkplus.org/. 2013 (cit. on p. 51).

[37] Software.intel.com. *Intel® Cilk™ Plus Reference Guide*. https://software.intel.com/en-us/node/522579. 2013 (cit. on p. 51).

[38] I. team (cit. on p. 11).

[39] *The polymath blog*. URL: http://polymathprojects.org/ (visited on 07/12/2010) (cit. on p. 56).

[40] *Trac Open Source Project*. (cit. on p. 57).

[41] *VirtualBox*. hrefhttp://virtualbox.orghttp://virtualbox.org. 2015 (cit. on pp. 42, 47).