History of changes

Grant agreement number 676541, June 2015

1. Updated participant acronyms for consistency with the EU portal.
2. Resources for UPSud: reinstated PhD position that was planned in the budget but went missing in the proposal document: +36PM for UPSud.
3. Reinstated related task T6.5 “Knowledge-based code infrastructure”. +33PM for WP6
4. Minor update to the involvement for Pons, Hivert, Lelièvre at UPSud for consistency with the submitted budget (-1PM each).
5. Minor update to the involvement for Gouarin at CNRS to make up for a higher salary than expected (-1PM); adapted accordingly WP2 (-1PM).
6. Where relevant: updated PM information according to the above.
7. WP2: Added mention of our participation to the European E-Infrastructure concertation activities.
8. Updated resources to be committed: audit costs should be in the direct costs, not subcontracting.
9. Updated the risk table to address the reviewers comments.
10. As suggested by the project officer, reduction of the number of deliverables (typically by merging together intermediate check-points into the corresponding final deliverable):
    - Suppressed irrelevant D1.1. (Consortium agreement).
    - Removed accidently duplicated deliverables:
      - D2.10 Course material on using OpenDreamKit in data science;
      - D2.12, D2.13 indexing service;
      - D2.20 Demonstrator: Interactive lecture notes and marking systems based on OpenDreamKit.
    - Merged together D1.5 and D1.8 (Data Management Plan V2,V3)
    - Merged D2.2, D2.7, D2.14, D2.19, D2.25, and D2.26 (community building reports: impact of development and training workshops) into a single yearly report.
    - Merged together D2.3, D2.15, D2.23, and D2.24 (Demonstrators: Problems in Physics with Sage v1,2 , Computational Mathematics for Engineering).
    - Merged together D2.8 (Community-curated indexing tool (open source)) and D2.9 (Community-curated indexing service for OpenDreamKit).
    - Merged D2.16 (Micromagnetic VRE code and documents source online), D3.4 (Python interface to OOMMF completed), D4.8 (Micromagnetic VRE software completed), D4.11 (Micromagnetic VRE tutorial and documentation notebooks), and D4.14 (Demonstrator online portal available) into a single deliverable D2.13 (Micromagnetic VRE completed and online).
    - Merged D3.1 and D3.9 (one-click install Sage distribution for Windows with Cygwin 32bit/64bit).
    - Merged D5.1 (Facility to compile PYTHRAN compliant user kernels and sage code) and D5.3 (Improve PYTHRAN runtime support to automatically take advantage of multi-cores and SIMD instruction units and use it in CYTHON); also switched lead info for D5.3 and D5.5 (Improve PYTHRAN typing to improve error information).
    - Merged D5.9 (Report on development of designs for the GAP developments – parallel library, interacts to standard infrastructure and CYTHON-like extensions) and D5.15 (Implementations of the GAP developments, ready for release) into D5.18 (Final report and evaluation of the GAP developments).
    - Merged D6.1 (DKS base survey and Requirements Workshop Report) and D6.3 (initial DKS base Design).
    - Merged D6.4 (Design of Triform (DKS) Theories (Specification/RNC Schema/Examples)) and D6.5 (Implementation of Triform Theories in the MMT API).
    - Merged D6.6 (LMFDB deep modelling: Fragment Identification and Initial Model Design), D6.7 (Heuristic Parser for the OEIS Import, Cross Validation of DKS-Model), D6.8 (Conversion of existing and new Databases to unified interoperable System).
    - Merged D6.10 (Full-text search (Formulae + Keywords) over Notebooks) and D6.12 (Formula search in CAS programs and Software Modules).
    - Merged D7.1, D7.4, D7.7 (Reports on relevant research in sociology of mathematics and lessons for design of OpenDreamKit VRE, parts I, II, III).

1st amendment of the Grant Agreement, June 2016

1. Addition of Ghent University (UGent) to the list of beneficiaries
2. Change of lead beneficiary from UPSud to UGent for D4.13
3. Addition of UGent the list of partners concerned by tasks: T1.1, T1.2, T1.3, T2.3, T3.1, T3.2, T3.3, T3.6, T3.7, T4.1, T4.4, T4.5, T4.12
4. Change of participation to WP1: 26.5 PM for UPSud and 1.5 PM fpr UGent
5. Change of participation to WP2: 9 PM for UPSud and 1 PM for UGent
6. Change of participation to WP3: 50 PM for UPSud and 14 PM for UGent
7. Change of participation to WP4: 12 PM for UPSud and 14 PM for UGent
8. Allocation of resources to UGent following their addition
9. Equivalent reduction of UPSud's resources following the addition of UGent
10. Reshuffling of resources allocated to JacobsUni following the organisation of an extra workshop (no change to their total budget)
11. Some deliverables are postponed, mostly due to difficulties in the recruitment process: D2.13, D3.2, D4.2, D5.2
12. Some tasks are postponed, mostly due to difficulties in the recruitment process: T2.7, T2.8, T2.9, T3.8, T4.11, T4.13, T4.14, T7.4
13. Stripped erroneously included pages about WP7 (page numbered 54 and 55 between 29 and 31); inserted missing page 30 about milestones.
14. Fixed minor typos in the texts.

2nd amendment of the Grant Agreement, March 2017

1. Accomodate the move of the KWARC Group (Prof Michael Kohlhase) from JacobsUni to FAU.
   (a) Addition of Friedrich Alexander University Erlangen/Nuremberg (FAU) to the list of beneficiaries.
   (b) WP1: moved 1 PM from JacobsUni to FAU.
   (c) WP6: Made FAU lead of WP6 (even though JU had been for the initial phase), moved 34 PM and the lead of all deliverables after month 15 from JacobsUni to FAU.
   (d) WP4: moved 8 PM and the lead of all deliverables after month 15 from JacobsUni to FAU.
   (e) Resources (Section 3.4): moved the respective resources from JacobsUni to FAU.
2. WP6: adapted the titles of D6.5, D6.7, and D6.8 to respect the change in priority on system interoperability and distributed computing in the "Math-in-the-Middle" Paradigm over algorithm verification.
3. Changed the lead of D3.9 from UOXF to UPSud, and the lead of D6.8 from UZH to FAU
4. Changed the lead of USFD from Neil Lawrence to Michael Croucher, and the breakdown of Person-Month to be used by this site (within the frame of the 360,00.00€ of their Personnel Costs)
5. Typo fix: D4.7 was meant to be due M24, not M14.
6. Changed the lead PI of UOXF from Ursula Martin to Dmitrii Pasechnik
7. Changed the lead PI of Simula from Hans-Peter Langtangen to Martin A. Sandves
8. UGA changed name from Universite Joseph Fourier to Universite Grenoble Alpes
9. UPSud changed its PhD position of 36 PM into a Postdoc position of 24 PM Reduced accordingly UPSud's involvement in PM's in WP6.
10. Consistency fix for UPSud: adjusted involvement in WP3 from 50 PM to 46 PM and reduced accordingly the PM's requested in UPSud's description for research engineers: upon the transfer to UGent, those figures had been incorrectly updated in the 1st amendment.
11. Accommodate the move of Hans Fangohr from SOUTHAMPTON to XFEL.
    (a) Addition of European XFEL GmbH/Schenefeld (XFEL) to the list of beneficiaries.
    (b) WP1: moved 1 PM from SOUTHAMPTON to XFEL.
    (c) WP2: Moved 15 PM and the lead of all tasks to be completed after month 24 and the deliverable D2.13 from SOUTHAMPTON to XFEL.
    (d) WP4: moved 5 PM from SOUTHAMPTON to XFEL.
    (e) WP7: moved deliverable D7.8 and 6 PM from SOUTHAMPTON to XFEL.
    (f) Resources (Section 3.4): moved the respective resources from SOUTHAMPTON to XFEL.
    (g) Removed 4,000 EUR for open access publication charges from Southampton (were by mistake omitted in submission and are thus not available).
12. Clarification of the CNRS site situation
    (a) Added Sébastien Labbé as a member of CNRS and removed 18 PM for the recruited engineer.
    (b) Clarification of PM repartition between CNRS and its third party University of Bordeaux Removed 28,198 EUR from "Other goods and services" from CNRS in order to take into account modifications in Person-Months involved in the project. This will have no impact on the deliverables in which CNRS is involved.
13. Postponment of deliverables within the first Reporting Period
    (a) D4.1 from M6 to M18
    (b) D4.3 from M12 to M18
    (c) D5.4 from M12 to M18
14. Some deliverables names in WP5 were modified
    (a) D5.5 from "Extend the existingassembly superoptimiser for AVX and upcoming Intel processor extensions for the MPIR library" to "Write an assembly superoptimiser supporting AVX and upcoming Intel processor extensions for the MPIR library and optimise MPIR for modern processors"

(b) D5.6 from"Parallelise the relation sieving component of the Quadratic Sieve and implement a parallel version of Block-Wiederman linear algebra over GF2 and the triple large prime variant" to "Parallelise the relation sieving component of the Quadratic Sieve and implement a parallel version of Block-Wiederman linear algebra over GF2 and implement large prime variants"

(c) D5.7 from "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" to "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 (e.g. AVX, etc.), especially in the FFT butterflies"

Work Plan revisions, Fall 2017

Following feedback from project officer and reviewers at the 18 month project review, we are proposing the following revisions:

1. Reduction of number of deliverables:
   - Merged D2.6 into D2.11.
   - Merged D2.8 into D2.9.
   - Merged together D2.5, D2.10 and D2.12 into D2.17
   - Merged D3.6 into D3.10.
   - Merged D4.14 into D4.12, to be delivered in M36.
   - Merged D5.8 into D5.11.
   - Merged D5.9 into D5.14.
   - Merged D5.10 into D5.16.
   - Merged D6.4 into D6.5.
   - Merged D6.11 into D6.5 and D6.8.
   - Merged D3.9 and D6.7 into D6.8.
   - Merged D6.6 into D6.10.
   - Merged D7.8 into D2.13.
2. Integration of WP7 activities into other WPs:
   - Canceled Objective 6 of OpenDreamKit
   - Integrated T7.4 in **T2.8**, lead by XFEL. Merged task is now **T2.8**.
   - Canceled **T7.1** and associated deliverable D7.5, and moved PM to WP3.
   - Canceled **T7.2** and associated deliverables D7.3 , D7.4 and D7.6 and moved PM to **T2.6**.
   - Canceled **T7.3**, cancelled deliverables D7.2 and D7.7, and moved PM to WP3 and WP5.

   The addition of manpower to WP3 and WP5 will enable an extension of the scope of the tasks **T3.1**, **T3.3**, and **T5.7**. Meanwhile, WP2 has proved to be very successful and OpenDreamKit is gaining momentum. The addition of manpower to this WorkPackage will allow the consortium to answer the increasing demand for collaboration and dissemination.
3. Work package specific milestones:
   - Introduced one milestone at month 42 for WP3, two milestones at months 36 and 48 for WP4, one milestone at month 36 for WP5 and two milestones at month 36 and 42 for WP6.
   - Removed redundant general milestone at month 36.
4. Complete rework of the Key Performance Indicators to better monitor the progress, impact, and results:
   - Four Indicators linked to the four project aims;
   - Focus on success stories and qualitative indicators in general;
   - Reusable, meaningful, and realistic quantitative indicators.
5. Miscellaneous
   - Canceled D3.4 as not relevant anymore: work has been done by the community. The corresponding workload has been reallocated, mostly to WorkPackage WP2 which needed more resources after the cancellation of WP7.
   - D4.11 moved from M24 to M36 to incorporate results from workshop on live structured documents.
   - **T4.2** expanded from M1-M24 to M1-M48 (the full duration of the grant) to reflect the ongoing nature of the work. Activity, total work, and deliverables are unchanged.
   - Typo fix in deliverable title for D1.5 (reports for year 2 and 3, not just year 2).
   - Modification of lead for WP6: from Jacobsuni to FAU
   - Remove the HPC conference in Bordeaux in favor of more dissemination schools and workshops (no modification of resources)
   - Reduction of 16 PM for UPSUD due to points 9 and 10 of the second Amendment. This modification had accidently not been implemented before in the part A of the Annex 1, therefore this update of the Annex 1 was necessary.
   - Termination of Jacobsuni on 17/01/2018: all relevant staff for the project was moved to FAU and Université Paris-Sud (originally planned on 31/12/2017 but was postponed to comply with the date of official submission of the amendment on the portal)
   - Termination of UZH on 17/01/2018: all activity planned for the partner has been accomplished and resources were spent (originally planned on 31/12/2017 but was postponed to comply with the date of official submission of the amendment on the portal)

Contents

1 Excellence

Improvements of the economy, ecology, health care, security and society overall are driven by innovation. 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 optimisation, 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 maximise 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 has many and varied aspects. To construct the OpenDreamKit virtual research environment toolkit we must consider component architecture, user interfaces, deployment frameworks and standardisation of software systems. We also study the social and technical questions needed to ensure the impact and sustainability of the virtual environment framework: its relationship to academic publication; open source tools and model development; data archiving and sharing and the 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: Maximise 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 standardise an architecture allowing combination of mathematical, data and software components with off-the-shelf computing infrastructure to produce specialised 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 Aims 3 and 4.

Objective 6: Explore the social phenomena that underpin collaboration in mathematics and mathematics software development (Canceled).

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 Aims 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 organisation, and training in the spirit of open-source software. This addresses Aim 4.

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 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

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

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.

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; standardisation 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). 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, virtualisation 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 specialised 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 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 specialised 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; modularisation; 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 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 organisation of "Sage-Days" all over the world. The first Sage-Days was held in 2006; to date there has 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 organisers 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

Below we explain how the project addresses the specific challenge and the scope of the topic "E-infrastructures for Virtual Research Environments (VRE)" under E-Infrastructures-2015 call, 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 and 9).
VRE should integrate resources across all layers of the e-infrastructure (networking, computing, data, software, user interfaces)	Our VRE will be assembled from OpenDreamKit components, which include mathematical *software* (from WP3 and WP5) *user interface* tools (from WP4) and *databases* (from WP6). In WP5 we will adapt them to interface to standard HPC and cloud environments, enabling *computing resources* to be included in a VRE. Specialised networking is not usually needed in this area.
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).
Each VRE should abstract from the underlying e-infrastructures using standardised 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 applications (Objectives 2 and 5). We will develop and demonstrate (WP2,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 optimisation 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 Millennium 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 desingularisation theorem of Hironaka, for which Hironaka won the Fields Medal, and its algorithmisation 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 including bulletin boards like `mathoverflow.net` and collaboration tools such as SAGEMATHCLOUD (see Section 1.3.9, page 16), Google Docs and `github`.

Open collaboration on Mathematical Software, Data, Knowledge In keeping with a strong tradition of sharing knowledge openly, Mathematicians started early on posting preprints on servers such as arxiv.org, publishing open online textbooks and monographs, contributing to Wikipedia, or building open source software. Due to the nature of the material, patents or restrictive copyrights are usually considered as irrelevant, when not as an obstacle to further innovation. Finally, mathematicians have a good record of credit recognition, with fights over intellectual ownership exceedingly rare.

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. One reason for this is that 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 1.3.1 on page 10 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. community-driven Q/A sites like [http://mathoverflow.net] with almost 40 thousand questions answered
   c. 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. formalisations of mathematical knowledge, e.g., in theorem prover libraries like Mizar [http://mizar.org], which has formalised 50 thousand relatively elementary theorems in 40 years or the formalisation 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 designer 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 standardised, which justifies the development of specialised 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 parallelisation and distribution of the computations is essential to cope with combinatorial explosion. Last but not least, they will need to visualise 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:

> Towards an Open Digital Research Environment 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.

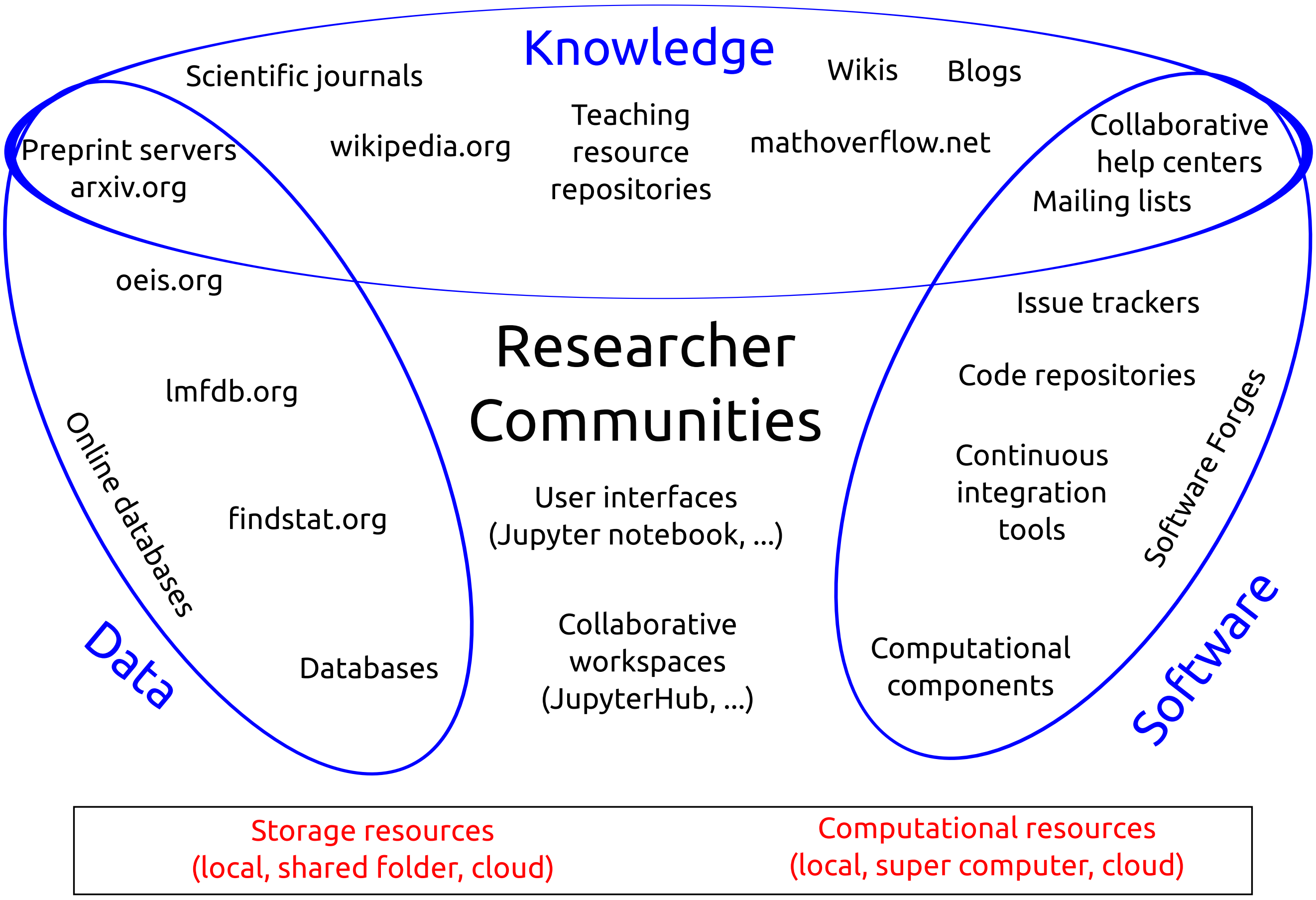

Figure 1.3.1: Virtual Research Environments for research in pure mathematics and applications.

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 maximise 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

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 16) 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 1.3.2 on page 11 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)

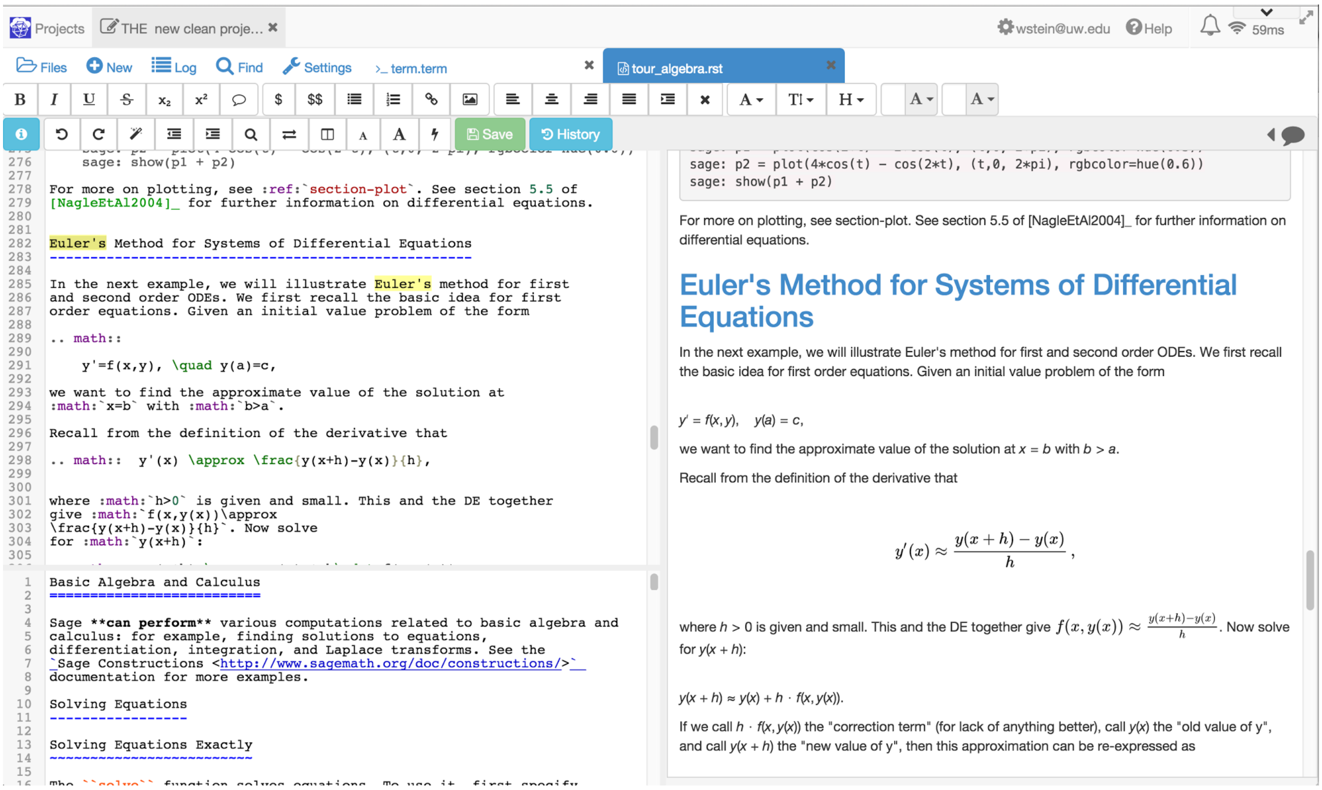

Figure 1.3.2: Typical SAGEMATHCLOUD session in a web browser

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 specialised 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 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 specialised 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.

A unifying user interface: executable notebooks and project Jupyter Project JUPYTER [Jup] is a set of open-source software projects for interactive and exploratory computing emerging from IPYTHON [Ipya]. 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 1.3.3 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 [Ket] and within [tea].

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.

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]

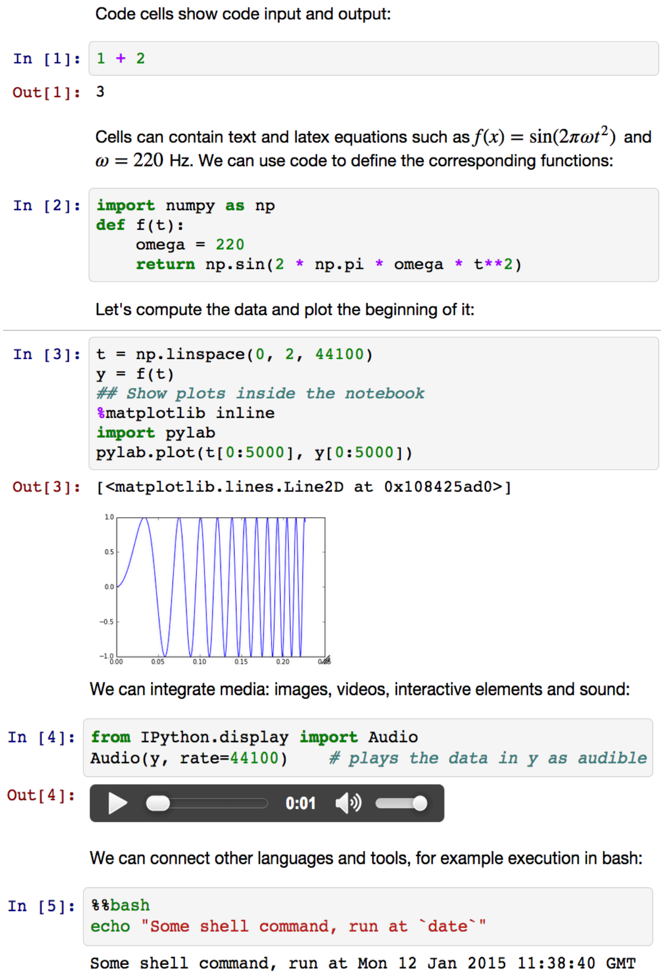

Figure 1.3.3: Self-contained JUPYTER Notebook demonstrating the concepts of cells that contain different types of material and can be executed (or updated) in arbitrary or sequential order.

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.

1.3.5 Approach

In previous sections, we have analysed 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 realise 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 decentralised 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

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

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 1 (WP1) covers project management and coordination as usual. 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;
- 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 and match up with the work packages defined in the implementation section: 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 WP2 is key to our goal to produce systems that address the end-user needs.

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 maximise 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 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.

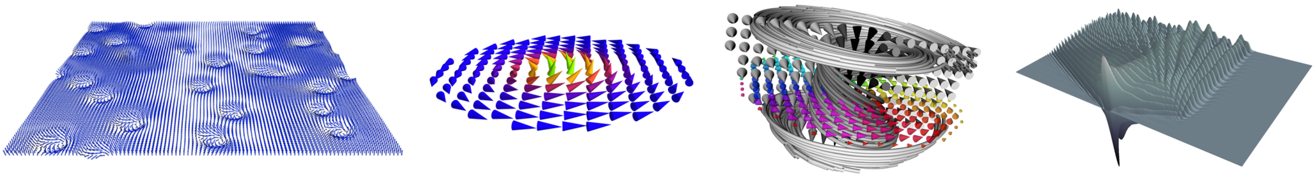

Figure 1.3.4: A selection of typical visualisation patterns often required in science and engineering. From left to right: a 3d vector field on a 2d domain, a 3d vector field coloured with another scalar field on a 2d domain, a 3d vectorfield on a 3d domain with streamlines, and scalar field plotted on a 2d domain.

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

To confirm the usability of our software tools, and demonstrate their value to potential academic and industrial users groups, we plan three demonstrator projects. These will be developed as soon as the necessary components are available to give early feedback and then refined and maintained during the remainder of the project to demonstrate the increasing scope and reliability of OpenDreamKit.

Demonstrator: Interactive Books — In **T2.9** we will produce a series of interactive books that can be read, executed, modified and explored within the OpenDreamKit VREs. This will demonstrate the collaboration and document-handling capabilities of OpenDreamKit and the ability to link presentation to computation. The books will target students, an audience familiar with modern web applications, but with traditional computational tools. Evaluation of the acceptance of these books with provide valuable feedback on our user interface components.

Demonstrator: Micromagnetic VRE In **T4.11** we will build a sample VRE targetting a specific research community, that of *micromagnetics*.

Micromagnetics deals with the behaviour of magnetisation 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. Figure 1.3.4 shows some micromagnetic simulation results. Researchers needing to carry out simulations in this area often do not have extensive computational background; they may be material scientists, engineers and physicists that interpreting experiments and designing new devices. Industrial users include Seagate, Hitachi, TDK, Samsung, Bosch and Toyota.

We will embed the most popular micromagnetic simulation software (Object Oriented MicroMagnetic Framework [Oom], which has over 1800 citations) within a micromagnetic VRE, complemented by other components that add value, including notebook user interfaces, replacing the current text driven, unautomated and error-prone workflow.

We will evaluate this demonstrator application in **T2.8**, 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.

Demonstrator: Computational Mathematics Resource Indexing Service In **T2.10** we will demonstrate the ability to build persistemnt highly-linked web-based services using OpenDreamKit by developing a unified index of components and documents relevant to OpenDreamKit. This will test elements such as cloud interaction and internationalisation, and our abvility to provide a uniform mathematical view of diverse components. It will also serve to advertise the range of components available.

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.

Thiéry and Stein, respectively coordinator and in the tentative advisory board of OpenDreamKit, are respectively foreign senior participant and PI of 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 parallelisation of exact linear algebra libraries and their composition, invent new algorithmic solutions for large scale parallelisations. 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 JacobsUni. 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 shell, 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 Encyclopedia 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.

1.4.1 Advances beyond the state of the art

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.

1.4.2 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.

1.4.3 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.

1.4.4 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 ambitious 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 Key Performance Indicators (KPI) show how OpenDreamKit addresses the specific impacts listed in the work programme. KPIs were thought through by the members of OpenDreamKit so that they are meaningful, reusable, realistic and easily measurable. The following qualitative and quantitative indicators are divided into the four aims of OpenDreamKit. If quantitative indicators are more useful for reporting and internal evaluation, qualitative indicators will give content for further dissemination and communication purposes, for example through the project website [5].
Aim 1: *Improve the productivity of researchers in pure mathematics and applications by promoting collaborations based on mathematical software, data, and knowledge.*

- Success stories reported as blogposts (Qualitative).

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.*

- Success stories about OpenDreamKit based VRE deployments and generally speaking adoption of OpenDreamKit's components (Qualitative);
- List of known OpenDreamKit based VRE deployments (Quantitative);
- Number of installs of OpenDreamKit's components via platform-specific distribution channels: Debian popcon, Arch statistics, installer downloads, etc. (Quantitative).

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.*

- Success stories (Qualitative);
- Number of PyPI hosted packages for SAGE, and similarly for other components (Quantitative);
- Number of additional systems made interoperable with the Math-in-the-Middle architecture, on top of the three for the Month 36 prototype (Quantitative);
- Metrics on the scale of the Math-in-the-Middle architecture; e.g. number of API CDs generated and number of alignments (Quantitative).

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

- Success stories resulting from dissemination activities such as workshops (Qualitative);
- Statistics on workshops organized and conference presentations delivered as part of our dissemination activities, including estimates of number of attendees (Quantitative);
- Number of courses and departments OpenDreamKit worked with directly and an estimate of how many students this subsequently affected (Quantitative).

2.1.2 Improving innovation capacity and the integration of new knowledge

Innovations developed by the 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 in 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:

[5] We will survey mathematical departments (and relevant members of other departments) at the end of each Reporting Period (M18, M36, M48) to gauge the awareness of the existence and capabilities of OpenDreamKit and its components, and to collect statistical data for estimating Key Performance Indicators listed in the table. The success factor is a positive change between the three surveys.

Market needs	How the project will address these needs
Performance gain	The toolkit will enable its users to combine functionality from several major open-source mathematical software systems and problem-oriented programming languages (including mainstream tools such as Python) in a modern integrated environment) 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 collaborate and work simultaneously on the same document, 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 from a wide pool of people 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, OpenDreamKit will enable new applications and features. All derived VREs are new applications, with new 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)

We start from SAGE's mission statement: "Creating a viable free open source alternative to MAGMA, MAPLE, MATHEMATICA and MATLAB" but need to go beyond that goal, and make OpenDreamKit a new reference tool, that is deployed across science. To be successful our VREs will need to provide much more than can be done through closed source commercial systems. Large scale collaborative development is the key to this goal, both on software and research, but coordination and projection of vision is still crucial for success. We will focus on the young generation, which will constitute its future users, for example by producing learning materials for the so called "Generation Y". "Generation Y" is expected to account for 30% of the total projected population in 2025, and will be the key influencer for 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 addressed and overcome using the mitigation strategies provided. These are distinct from the risks to project delivery detailed in Section 3.2.5.

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* have 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 valuable advice on real user needs throughout the project 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.

Table 2.1: Barriers to Impact

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 dissemination and exploitation 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.

We summarise the dissemination activities and how they will help to achieve the expected impact among our stakeholders and target audiences in Table 2.2.2.

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;
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);
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 VREs promoted by OpenDreamKit will consist of a thin layer on top of an ecosystem of open source components. Hence, the long term sustainability of those VREs is guaranteed by the sustainability of the ecosystem of components (Objective 5).

By the end of the project, we expect that the main barriers will have been addressed, and that the needs for financial support will therefore not be very important. Furthermore, we expect that some of developers' positions will be extended beyond the duration of the project by the partners' institutions, due to the 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 companies will desire to use and further improve and benefit from OpenDreamKit VREs or components thereof. This will lead to further feedback and improvements, and also will open the door for the attracting additional funding, either:

- through access provision to other scientific communities, on projects base, or via service delivery; this opportunity is for example already being explored by the SAGEMATHCLOUD project in the US: it recently spun off a company on this business model to seek for additional funding.

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 Standardisation agencies;
- T6 Civil society;
- T7 Public at large.

Table 2.2.2: Dissemination and exploitation plan

- through specific developments and training services delivered by companies like Logilab that have a solid business model based on open source software.

We conclude by a quote of Stéphane Fermigier, president of the Open Source Software Working Group[6] of the Systematic Paris Region Systems & ICT Cluster (having UPSud and Logilab as members), and founder of NUXEO[7]:

> Open source is still reshaping all aspects of the software industry, specially high growth sectors such as Big Data, Enterprise 2.0 or Mobile Applications. NOT using open source components is now considered the exception rather than the rule in almost all companies that produce software, creating a tremendous opportunity for the Paris Region open source ecosystem, whose leadership has been recognised for a long time, from academic research teams to young innovative software vendors, from specialised consultancies to large systems integrators.

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 1.3.1 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.

[6] `http://www.systematic-paris-region.org/en/get-info-topics/free-and-open-source-software`
[7] http://www.nuxeo.com/

3 Implementation

3.1 Work Plan — Work packages, deliverables and milestones

3.1.1 Overall Structure of the Work Plan

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 27 illustrates the timeline for the various tasks for these 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 standardise 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	Cancelled
Objective 7	Identify and extend ontologies	WP6
Objective 8	Effectiveness of the VRE	WP2
Objective 9	Effective dissemination	WP2

Work Programme for Objective 1: **T3.2** (Interfaces between Systems) directly addresses the core of objective 1, making existing systems compatible with one another in mathematically sound ways. Other tasks in WP3 (component architecture) support this, by making components more portable and easier to deploy (**T3.3**: Modularisation and Packaging; **T3.1**). **T3.6** will bring us the benefit of lessons learned and components built for SAGEMATHCLOUD. **T6.3** deals with the data-centric aspects of the interfaces. Additionally elements of WP4 (user interface) and WP5 (HPC) will also contribute to the framework with user interface pluggability and interfaces optimised for HPC.

Work Programme for Objective 2: We have identified a need for a number of new core components for OpenDreamKit and planned their construction at appropriate stages of various workpackages. A new adapter infrastructure is part of **T3.2**; new virtual appliances will be built in **T3.1**; new components will be extracted from SAGEMATHCLOUD in **T3.6**; new documentation components will be developed in **T4.4** and **T4.5**; new mathematical software will be developed in **T5.6** and new database tools in **T6.9** and **T6.10**.

Work Programme for Objective 3: Representatives of a number of communities have already come together simply to prepare this proposal, and the whole project will work to bring them together. Specifically developers of many systems will be brought together to complete work package WP3, especially **T3.2**. Bringing broader communities together is the core purpose of work package WP2, which includes workshops, web sites, demonstrator packages and outreach activities.

Work Programme for Objective 4: The concept of this project is centred on leveraging the communities vast investment in existign open source software systems, and wherever possible we will proceed by extending and updating existing software components. In work package WP3 we will address portability (**T3.1** and modularity (**T3.6**) and also adapt the components to use the new interfaces being designed in **T3.2**. Work package WP5 is largely about updating software for performance, while workpackage WP4 deals with adaptation of UI components and of other systems to work with them.

Work Programme for Objective 5: A number of tasks relate to developing promoting and supporting sustainable models for collaborative software development. On a practical level **T3.7** will adress processes and technologies, **T6.9** concerns collaborative accumulation of data. On a personal level, much of WP2 deals with ensuring a wide and committed user/developer community.

Work Programme for Objective 6: Objective 6 was covered by a dedicated work package WP7 on social aspects. It ranged from analysis of the needs with **T7.1** to evaluation with **T7.3**. That objective and dedicated WP were cancelled.

Work Programme for Objective 7: Objective 7 is addressed directly by WP6, which deals with data and its meaning.

Work Programme for Objective 8: Producing and evaluating systems that demonstrate our achievements is a key feature of this project, and this work in embedded throughout the project. The integration and publicisation of these demonstrators is key to WP2 (dissemination) especially later in the project, while their formal evaluation is found in WP7.

Work Programme for Objective 9: Dissemination is the heart of WP2. Members of OpenDreamKit will organise workshops within **T2.5** or **T2.6** as well as less formal meetings with interested groups. In addition, we will follow open software development processes throughout the project, so that our work is immediately available to any interested party. We will announce important developments or releases through our own web pages and the established channels of the component systems. Our scientific findings will be published in the open scientific literature and announced at scientific meetings and conferences in the usual way and reported in annual project reports.

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 tasks are fairly independent from each other, which is reflected by the GANTT chart in Figure 3.1.1

Figure 3.1.1: Gantt Chart: Overview Work Package Activities – *Bars shown at reduced height (e.g. 50%) indicate reduced intensity during that work phase (e.g. to 50%).*

3.1.4 Milestones

By nature and design the project consists of a large number of loosely coupled tasks, especially so in the first stages. We have therefore chosen to schedule two first milestones that track general progress on the project. Later milestones track and showcase how the achievements in certain collections of tasks in the work packages come together to produce high level user-visible features, in particular in the context of the Virtual Research Environment.

The milestones have been scheduled at the occasion of the project yearly meetings, where they can be discussed in detail, tracking the progress in each work package through status reports on the tasks and deliverables and take corrective measures, where necessary, and critical decisions regarding further plans. The later milestones coincide as well with the formal project reviews for demonstration, assessment and discussion with the reviewers. We envisage that this setup will give the project the vital coherence in spite of the broad interdisciplinary mix of various backgrounds of the participants.

General Milestones

1. **Milestone M1 (Month 12) Startup** By this milestone we will have carried out the requirements study, design and prototype implementations and started community building activities.

2. **Milestone M2 (Month 24) Implementations** By this milestone we will have constructed first fully functional interface implementations and released enhanced versions of OpenDreamKit components, and train early adopters of OpenDreamKit.

3. **Milestone M4 (Month 48) Evaluation** By this milestone we will have released final versions of all OpenDreamKit components and completed the project evaluation.

Milestone for WP 3

1. **Milestone M5 (Month 42) ODK's computational components available on major platforms** User story: users shall be able to easily install ODK's computational components on the three major platforms (Windows, Mac, Linux) via their standard distribution channels.

Milestones for WP4

1. **Milestone M6 (Month 36) Prototype VRE for mathematical researchers** User story: a group of mathematical researchers with access to common computational resources, such as a shared lab computer or cloud servers, shall be able to deploy a prototype VRE with JUPYTERHUB, integrating OpenDreamKit components. The Jupyter kernels for mathematical software developed as part of OpenDreamKit make computational mathematical components accessible in a JUPYTER environment, enabling a Jupyter-based deployment of the relevant tools for the researchers. The process of working on notebooks is greatly improved by review tools developed as part of WP4, enabling researchers to collaborate to some degree in a shared computational environment.

2. **Milestone M7 (Month 48) Collaborative VRE for mathematical researchers and beyond** The prototype VRE shall be extended with improved ease of deployment, new functionality such as interactive 3D visualization and real-time collaboration, enabling researchers to collaborate productively in a shared computational environment. Finally, integrating notebooks and semantic knowledge into a publication / knowledge system enable a continuous process of leveraging OpenDreamKit components from research to publication.

Milestone for WP 5

1. **Milestone M8 (Month 36) Seamless use of parallel computing architecture in the VRE (proof of concept)** User story: Astrid wants to run compute intensive routines involving both dense linear algebra and combinatorics. She has access through a JupyterHub-based VRE to a high end multi-core machine which includes a vanilla SAGE installation. She automatically benefits from the HPC features of the underlying specialized libraries (LINBOX, ...). This is a proof of concept of the overall framework to integrate the HPC advances of specialized libraries into a general purpose VRE. It will prepare the final integration of a broader set of such parallel features for the end of the project.

Milestones for WP 6

1. **Milestone M9 (Month 36) First Math-In-The-Middle-based interoperability prototype**

   User story: thanks to a fully functional prototype integrating of at least the systems GAP, SAGE, SINGULAR, and LMFDB via the SCSCP Protocol, end users shall be able to run calculations involving any combination of those systems from any of them. This prototype will be the basis for integration work for additional systems and the user interface from WP4.

2. **Milestone M10 (Month 42) Second Math-In-The-Middle-based interoperability prototype** The goal of this milestone is to take into account all the operational experiences with the first prototype and add more systems and integrate some of the UI components from WP4. The experiences with the preparation of this prototype will allow us to estimate the joining costs of adding a system to the OpenDreamKit VRE toolkit, which is an important measure of the flexibility of the Math-In-the-Middle approach.

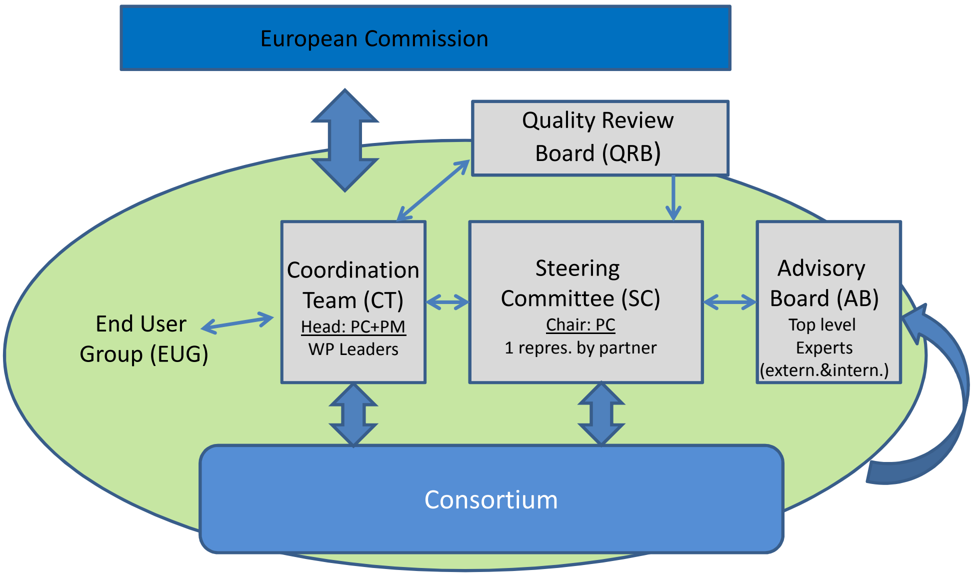

Figure 3.2.1: Management structure

3.2 Management Structure and Procedures

3.2.1 Management

The project will be coordinated by the Université Paris-Sud (UPSud), represented by Prof. Nicolas M. Thiéry (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 organised 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 UPSud. Additional feedback and expertise will be brought by Financial, Legal and European affairs officers from SAIC.

3.2.2 Organisational structure and decision-making

The organisational structure, shown in the Figure 3.2.1, has been designed to enable efficient coordination of the 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 mobilise 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 mobilised 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 organisational 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 organisation.

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 organisations, including both experts from the project scientific area, and experts on legal and social matters. The following external persons have already confirmed their interest: William Stein[11], lead developer of SAGEMATHCLOUD, one representative of the Open Source Software Working Group[12] of the Systematic Paris Region Systems & ICT Cluster, and Istvan Csabai[13] of the Wigner Research Centre for Physics (WRCP) of the Hungarian Academy of Sciences.

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.

[11] wstein.org
[12] http://www.systematic-paris-region.org/en/get-info-topics/free-and-open-source-software
[13] http://complex.elte.hu/~csabai/

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 organise 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), specialised 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.

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.

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.

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, Versailles), SAGE (Orsay, Versailles, CNRS, 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 optimisation/parallelisation (Logilab, Grenoble)
7. A company specialised 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 3.1.5, 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 SAGEMATH-CLOUD 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 optimisation 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.

	UPSud	CNRS	JacobsUni	UGA	UNIKL	UOXF	USlaski	USFD	SOUTHAMPTON	USTAN	UVSQ	UWarwick	UZH	Logilab	Simula	UGent	FAU	XFEL
UPSud		@		@	@	@				@	@	@@	@@	@		@		
CNRS	@			@	@●	@				@	@	@@●	@●			@@		
JacobsUni					☺					○○			○				☺○○○	
UGA	@	@			@	@				@○	@	@	@			@		
UNIKL	@	@●	☺	@		@@				@@●○★	@	@●	@●			@	☺	
UOXF	@	@		@	@@				●	@@●★	@	@●	@			@		●
USlaski															●			
USFD									@						@	@		@
SOUTHAMPTON						●		@							@	@		@●
USTAN	@	@	○○	@○	@@●○★	@@●★					@	@●	@			@	○○	
UVSQ	@	@		@	@	@				@		@	@			@		
UWarwick	@@	@@●		@	@●	@●				@●	@		@@●	@		@@		
UZH	@@	@●	○	@	@●	@				@	@	@@●		@		@	○	
Logilab	@											@	@					
Simula							●	@	@							@		@
UGent	@	@@		@	@	@		@	@	@	@	@@	@		@			@
FAU			☺○○○		☺					○○			○					
XFEL						●		@	@●						@	@		
joint	★≙ publication, ●≙ project, ○≙ organization, @≙ software/resource dev, ☺≙ supervision																	

Table 3.3.1: Previous Collaboration between OpenDreamKit members

3.4 Resources to be Committed

3.4.1 Management Level Description of Resources and Budget

Staff efforts By design OpenDreamKit is attacking upfront the diverse needs of a very large community: pure maths and applications. Thanks to its toolkit approach, it will impact users in many areas of science. This is a major long term investment which requires the extensive expertise of a large consortium and the improvement of a great number of software components. The complex technical nature of the project, combined with the high quality requirements for guaranteeing long term sustainability, necessitates recruiting highly experienced software engineers to complement the participants. Much emphasis is put as well on studying the social aspects and implementing many demonstrators to illustrate the breath and depth of potential applications.

This all explains the considerable staff efforts.

Travel, dissemination, and outreach 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.

Guidelines for travel and 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 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.

Guidelines for 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.

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 UPSud requests 12 person months for the project coordinator (Nicolas M. Thiéry), 5 person months for two researchers (Florent Hivert, and Samuel Lelièvre) and for the lead PI (Viviane Pons), 44 person-months for a full time developer and 5.5 person-months for a part-time developer, 24 months for a postdoc, and 24 months for a part time project manager for the full duration of the project.

1: UPSud	Cost (€)	Justification
Travel	85,139	Travel (see the guidelines 3.4.1)
Publication charges	1,000	Open access publication charges (see 3.4.1)
Other goods and services	92,000	8 developer workshops **T2.3**, 4 training and dissemination workshops **T2.1**, audits certificates on the financial statements
Total	178,139	

Table 3.4.1: Overview: Non-staff resources to be committed at UPSud (all in €)

Resources CNRS CNRS requests 14 person months for the lead PI Vincent Delecroix, 13 for the PARI/GP head Karim Belabas, 8 person months for PIs Adrien Boussicault, 7 person months for PI Bill Allombert, 6 person months for PI Sébastien Labbé, 5 person months for PI Loïc Gouarin, and 30 person months for a full time developer working on tasks **T5.1**, **T5.6**, **T4.10** and **T4.12**.

2: CNRS	Cost (€)	Justification
Travel	56,700	Travel (see the guidelines 3.4.1)
Publication charges	1,000	Open access publication charges (see 3.4.1)
Other goods and services	82,802	4 Ateliers PARI **T2.3**, A series of dissemination workshops and trainings in developed and developing countries **T2.5**
Total	140,502	

Table 3.4.2: Overview: Non-staff resources to be committed at CNRS (all in €)

Resources Jacobs University Bremen JacobsUni requests 6 PM each for a research programmer (Dr. Christian Maeder) and a junior researcher (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. The rest of the PM (11) are to be dispatched within the other JacobsUni participants.

3: JacobsUni	Cost (€)	Justification
Travel	8,000	Travel (see the guidelines 3.4.1)
Publication charges	1,500	Open access publication charges (see 3.4.1)
Equipment		
Other goods and services		
Total	9,500	

Table 3.4.3: Overview: Non-staff resources to be committed at JacobsUni (all in €)

Resources Université Grenoble Alpes UGA requests 12 person months for an engineer (Pierrick Brunet) starting in month 0 and working on **T5.7**, 24 person months for another engineer starting in month 12 and working on **T5.3**, 15 person months for the lead PI (Clément Pernet) and 9 person months for a 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.

4: UGA	Cost (€)	Justification
Travel	60,850	Travel (see the guidelines 3.4.1)
Publication charges	1,000	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	34,000	2 developer workshops: HPC and Pythran, audits certificates on the financial statements
Total	119,850	

Table 3.4.4: Overview: Non-staff resources to be committed at UGA (all in €)

Resources University of Kaiserslautern UNIKL requests 48 person months for a researcher to work on tasks **T5.4** (46 PM), **T4.1** (2 PM), 12 person months for a researcher starting in month 6 to work on **T5.5**. The cost for Professor Decker's activities within OpenDreamKit (6 PM), including the related overhead, will be covered by UNIKL and is therefore not part of the requested funding. The lead PI will take on all management responsibilities.

5: UNIKL	Cost (€)	Justification
Travel	67,100	Travel (see the guidelines 3.4.1)
Other goods and services	67,600	5 developer workshops **T2.3**, audits certificates on the financial statements
Total	134,700	

Table 3.4.5: Overview: Non-staff resources to be committed at UNIKL (all in €)

Resources University of Oxford University of Oxford requests in total 29 Person-Months for its lead PI Dmitrii Pasechnik, Ursula martin and Edith Elkind. 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.

6: UOXF	Cost (€)	Justification
Travel	25,000	Travel (see the guidelines 3.4.1)
Equipment	3,000	Laptop and a large display for a co investigator
Total	28,000	

Table 3.4.6: Overview: Non-staff resources to be committed at UOXF (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 visualisation 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 visualisation.

The justification for subcontracting is as follows. We have much experience as users of 3d visualisation 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.

7: USlaski	Cost (€)	Justification
Travel	18,188	Travel (see the guidelines 3.4.1)
Total	18,188	

Table 3.4.7: Overview: Non-staff resources to be committed at USlaski (all in €)

Resources University of Sheffield Sheffield requests 60 person months to be split between two researchers, 9 person months for the lead PI (Mike Croucher) and 2 person months for the initial leadPI (Neil Lawrence). The lead PI will take on all management responsibilities. Two researchers will be employed, one with a specific focus on the SGE Implementation **T5.8**. The other for a period of whose focus will be on user interfaces, social aspects **T7.2** and course development **T2.6**. Researchers 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 (D7.4) at 5,000. For the researchers we request screens and computers (3,000).

8: USFD	Cost (€)	Justification
Travel	53,200	Travel (see the guidelines 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	28,000	Workshops (Project meeting and two dissemination workshops, see **T2.6**), HPC Compute Time (see **T5.8**), audits certificates on the financial statements
Total	95,200	

Table 3.4.8: Overview: Non-staff resources to be committed at USFD (all in €)

Resources Southampton Southampton requests 16 person months for a researcher and 3 person months for the lead PI (Hans Fangohr). The lead PI will take on all management responsibilities.

9: SOUTHAMPTON	Cost (€)	Justification
Travel	15000	Travel (see the guidelines 3.4.1)
Equipment	4,000	two high performance laptops
Total	19,000	

Table 3.4.9: 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.

10: USTAN	Cost (€)	Justification
Travel	76,800	Travel (see the guidelines 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	21,500	2 dissemination workshops (room hire and subsistence for external participants; task **T2.3**), audits certificates on the financial statements
Total	117,300	

Table 3.4.10: Overview: Non-staff resources to be committed at USTAN (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 UPSud, Université de Versailles is not going to hire any full-time personnel for the project.

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

Table 3.4.11: Overview: Non-staff resources to be committed at UVSQ (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.

12: UWarwick	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 3.4.12: Overview: Non-staff resources to be committed at UWarwick (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**, **T6.7** and **T6.8**. He is in charge of deliverables D6.4, and D6.8.

13: UZH	Cost (€)	Justification
Travel	26,800	Travel (see the guidelines 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 3.4.13: Overview: Non-staff resources to be committed at UZH (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**.

14: Logilab	Cost (€)	Justification
Travel	17,200	Travel (see the guidelines 3.4.1)
Total	17,200	

Table 3.4.14: Overview: Non-staff resources to be committed at Logilab (all in €)

Simula Research Laboratory Simula requests 29 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.

15: Simula	Cost (€)	Justification
Travel	56,200	Travel (see the guidelines 3.4.1)
Publication charges	4,000	Open access publication charges (see 3.4.1)
Other goods and services	11,500	Organisation of the Jupyter workshop, audits certificates on the financial statements
Total	71,700	

Table 3.4.15: Overview: Non-staff resources to be committed at Simula (all in €)

Resources Ghent University Ghent will employ one person associated with the project, Jeroen Demeyer, who is also the PI for UGent. 14 person-months will be dedicated to WP3 and 14 person-months to WP4. He will additionnaly use a 1.5 person-month for WP1 and 1 person-month for WP2. He is in charge of D4.13.

16: UGent	Cost (€)	Justification
Travel	10,000	Travel (see the guidelines 3.4.1)
Total	10,000	

Table 3.4.16: Overview: Non-staff resources to be committed at UGhent (all in €)

Resources European XFEL GmbH XFEL requests 24 person months for a researcher, and 3 person months for the lead PI (Hans Fangohr), to lead and complete the work started at Southampton.

17: XFEL	Cost (€)	Justification
Travel	36,100	Travel (see the guidelines 3.4.1)
Equipment	6,000	HPC Workstation (6k) to host micromagnetic VRE web server, **T4.14**
Other goods and services	15,200	Dissemination workshops, audits certificates on the financial statements
Total	57,300	

Table 3.4.17: Overview: Non-staff resources to be committed at XFEL (all in €)

Resources FAU Erlangen Nürnberg FAU 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 12 PM each for 2 junior researchers. One will do much of the actually system development in WP4 and WP6 while the other will concentrate on the cases studies in WP6. The remaining PM (7) are to be dispacthed between the FAU participants according to the Table 3.4.1.

18: FAU	Cost (€)	Justification
Travel	42,100	Travel (see the guidelines 3.4.1)
Publication charges	2,700	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 **T6.10**
Other goods and services	7,600	Audits certificates on the financial statements and organisation of one workshop/interim review
Total	66,400	

Table 3.4.18: Overview: Non-staff resources to be committed at JacobsUni (all in €)

4 Members of the Consortium

4.1 Participants

4.1.1 UPSud: 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 or contribute to tasks on or around the SAGE computational system (**T3.6**, **T3.2**, **T4.1**, **T4.4**, **T4.5**, **T5.6**, **T6.9**). Finally, it will lead WP2 and in particular host or coorganise many of the community building, training, and dissemination actions.

Curriculum vitae of the investigators

Viviane Pons (leadPI, female, 5 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, 5 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[14]. 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, 5 PM) Maître de conférences since 2006 at Laboratoire de mathématique d'Orsay, Université Paris-Sud, Samuel Lelièvre is an established researcher in Dynamics and Geometry, with 10 papers published in international journals including Annales scientifiques de l'École normale supérieure, Crelle, GAFA, Geometry and Topology. He participated in three ANR projects, and has collaborators in France, Israel, the UK, the USA. His research in Dynamics and Geometry often involves explicit and experimental approaches, for which he writes code in order to explore combinatorial objects such as square-tiled surfaces, translation surfaces, group actions, group presentations.

He uses and actively promotes SAGE since 2010. He is in the top 15 contributors of the 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

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

tutorials twice at the GDR-IM yearly school for French PhD students at the interface of Mathematics and Computer Science, and at the CIMPA/ICPAM school Bobo2012 on Discrete Mathematics (Bobo Dioulasso, Burkina Faso, 2012).

Loïc Gouarin (R, male, 0 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 administrative constraints within CNRS, Loïc Gouarin will be attached to CNRS.

NN (res, 50 PM) We will hire two full time experienced software developers to work under the leadership of Nicolas M. Thiéry on the technical tasks pursued by UPSud, 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 a postdoc in computer science with a strong background in mathematics, to conduct research within WP3, WP4, and WP6, and in particular on **T6.5**.

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.

To make this into a more attractive full time position, we are planning a joint hire with another European project lead by CNRS at a nearby institution.

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

UPSud 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.

UPSud 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 CNRS: CNRS (FR)

The Centre National de la Recherche Scientifique (National Centre for Scientific Research, CNRS) is a public organization under the responsibility of the French Ministry of Education and Research. Its missions are

- To evaluate and carry out research and bringing social, cultural, and economic benefits for society.
- To contribute to the application and promotion of research results.
- To develop scientific information.
- To support research training.
- To participate in the analysis of the national and international scientific climate in order to develop a national policy.

It consists of ten institutes in Biology, Humanities, Information Sciences and Technologies, etc. CNRS laboratories (or research units) are located throughout France, and employ a large body of tenured researchers, engineers, and support staff.

Three research units are concerned by this proposal. The Laboratoire Mathématiques d'Orsay (LMO) which is a Joint Research Unit with Université Paris-Sud which is a partner of this proposal and has already been described (see 4.1.1). Two laboratories in Bordeaux (FR), the Laboratoire d'Informatique Bordelais (LaBRI) and the Institut Mathématiques de Bordeaux (IMB). Those two units are joint with the Université de Bordeaux which is a Third Party Linked to CNRS of this proposal (see 4.2).

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. The former is a leader in Europe for analytic combinatorics.

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 Bordeaux site will lead the development of PARI/GP within workpackages WP4,WP5, WP2 and will play a major role in WP5 with respect to combinatorics in partnership with UPSud. It will also organise workshops in favour of developing countries (WP2).

Due to lack of excellent candidates, the non-permanent engineer was recruited on month 18 of the project. As a consequence, he will be involved for 30PM on the project instead of the 48 initially planed. To compensate this lost extra PM has been attributed to permanent researchers involved in the projet and the deliverables will be provided on time.

Curriculum vitae

Vincent Delecroix (leadPI, male, 14 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, 13 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, 7 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, 8 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.

Sébastien Labbé (R, male, 6 PM) CNRS researcher at the LaBRI (Bordeaux, France) from january 2017, Sébastien Labbé is a junior researcher in Word Combinatorics, Discrete Geometry and Dynamical Systems. He published 15 articles in international journals and 15 articles in conferences.

Since 2008 he is an important contributor to SAGE with more than 100 tickets authored and reviewed. He publishes its research code yearly as a Sage Optional package containing more than 20000 lines of code. He authored many blog posts on `planet.sagemath.org`. He also organised many Sage days and Sage workshops in Canada, France and Belgium.

NN (R, 30 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:

- creation of CYTHON/PYTHON bindings for PARI/GP and more generally Sage packaging,
- 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.).

Loïc Gouarin (R, male, 5 PM) Loïc Gouarin is working at Université Paris-Sud (see 4.1.1 for his CV) but is a CNRS employee and will thus be administratively attached to Bordeaux.

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).

4.1.3 JacobsUni: 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 [KWARC]) 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.

Prof. Dr. Michael Kohlhase has moved to FAU Erlangen/Nürnbert 1. September 2016, but retains an adjunct professorship at Jacobs University until 31. August 2017.

In total JacobsUni requests 17 Person-Months, the 4 PM for WP4 and 12 PM for WP6 to be dispactched between the members of the KWARC team.

Curriculum vitae

Michael Kohlhase (leadPI, male, 0 PM) See FAU

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, 5 PM) See FAU

Mihnea Iancu (JRes, male, 6 PM) See FAU

Relevant previous experience:

See FAU

Specific expertise:

See FAU

4.1.4 UGA: Université Grenoble Alpes (FR)

The Université Grenoble Alpes (UGA), previously known as 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. UGA laboratories have successfully contributed to the FP7 european with 105 projects, 26 of them being coordinated by UGA. Under H2020, UGA continues to actively participate with already 4 projects.

The UGA 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 **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 UNIKL: 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, whose lead developer, William Heart, is working at UK.

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.

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 Wolfram Decker 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.

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.

4.1.6 UOXF: 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

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.

Edith Elkind (PI, female, 3 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.

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 that was to be undertaken in the Work Package 7 (Social Aspects) fitted very well into her current project, which concerned crowdsourced mathematics: the overarching goal was to understand and extend the human and computer creation of mathematics. It was mostly funded by her 2014 EPSRC Advanced Fellowship (EPSRC awards only one or two of these annually in Computer Science) which is a partnership of industry, government and international academia.

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/S00355ED
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 USlaski: 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.

4.1.8 USFD: 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.

Dr Mike Croucher is a leading expert in Research Software Engineering (RSE) based in the Department of Computer Science. He has an EPSRC RSE fellowship and alongside Dr Paul Richmond (also an RSE Fellow) he leads the University's efforts in Research Software Engineering. Mike is a Fellow of the Software Sustainability Institute and has over 10 years experience in Research Software at the University of Sheffield and the University of Manchester.

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. Most of the research tasks will be accomplished by the two researchers hired on the project, under the supervision of Mike Croucher who has replaced Neil lawrence as leadPI in the course of the project. The total amount ot PM goes up to 72.

Curriculum vitae

Michael Croucher (leadPI, male, 9 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.

Neil Lawrence (PI, male, 2 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[15] 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.[16]

NN (R, 30 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.

[15] `https://github.com/SheffieldML/GPy`
[16] `http://ml.dcs.shef.ac.uk/gpss/`

NN (R, 30 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/).

4.1.9 SOUTHAMPTON: 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.

Prof. Hans Fangohr and his group move to European XFEL GmbH as of 1. September 2017, and will complete all OpenDreamKit related tasks initially scheduled for Southampton there. After the move and completion of associated administrative work, Southampton will not be involved in the OpenDreamKit work programme anymore.

Curriculum vitae

Hans Fangohr (leadPI, male, 3 PM) Hans Fangohr is Professor of Computational Modelling at the University of Southampton until August 2017 (3PM), then Senior Data Analysis Scientist at the European XFEL GmbH (3PM). 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`) at Southampton, 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) 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.

Marijan Beg (R, male, 16 PM) Dr Marijan Beg is a post-doctoral researcher with experience in computational science, IPython and the Jupyter Notebook and micromagnetic simulations. He is working under the leadership of and together with Hans Fangohr at Southampton until August 2017 (16PM), and then moving to continue the work at European XFEL GmbH from September 2017 onwards (24PM).

Publications, products, achievements

See European XFEL GmbH (page 64).

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)

4.1.10 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 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 GPL, and our records show that it has been installed in at least 3000 sites and cited in several thousand 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 and Fort Collins.

CIRCA publishes highly-regarded research in Pure Mathematics and in Computer Science, recognised by our highly-cited publications in top international venues in both disciplines. A particular speciality is research that integrates the two disciplines very closely, for instance the study of decidability problems in algebraic structures defined using formal language theory. 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[17]. 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/).
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.

[17] Applicable Algebra and Error Correcting Codes

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 continuous integration systems used by GAP and its large collection of packages, and manages GAP release preparation. We have a substantial pool of multicore compute servers.

4.1.11 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.

4.1.12 UWarwick: 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.

4.1.13 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 M. Thiéry. 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).

4.1.14 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 visualisation 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 visualisation 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.

4.1.15 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

Martin Sandve Alnæs (leadPI, male, 8 PM) Martin Sandve Alnæs is project leader for the Computational Middleware project at 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.

Alnæs received his PhD from the Department of Informatics, University of Oslo, in 2009, and then worked as Senior Software Developer at TANDBERG and Cisco before being hired as a Postdoctoral Fellow in the Department for Biomedical Computing at Simula in 2011. He is now a Senior Research Engineer at Simula since 2015.

Alnæs' research involves novel approaches to scientific software design as well as application of computational fluid dynamics to blood flow in aneurysms. Alnæs' is one of the main developers of the FEniCS software suite for automation of finite element simulations, where his main contributions have been on the use of domain specific languages, symbolic computing, and code generation, to achieve high productivity and high performance in a general framework.

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.

4.1.16 UGent: GHENT UNIVERSITY (BE)

Today Ghent University attracts over 41000 students, with a foreign student population of about 4100. In 2015, the university invested over 242 million Euros in research projects on behalf of public and private partners, and employed around 6600 academic staff members. Ghent University is ranked 71st in the Shanghai and 118th in the Times ranking. The University has participated in more than 200 research projects in the EU's Sixth Framework Programme (2002–2006) and in 260 projects in the Seventh Framework Programme, of which 27 ERC grants and 26 Marie Curie Fellowships. Ghent University coordinated 42 collaborative projects in FP7. Up till now, Ghent University is under H2020 (the new Framework Programme (2014–2020)) involved in 100 projects, and coordinates 9 of them. Among these newly acquired projects, Ghent University hosts 15 ERC grant holders and supervises 9 Marie Skłodowska Curie fellowships. The university provides excellent training opportunities to both young and experienced researchers, and awarded 660 PhD degrees in 2015 of which over 30% went to international young researchers. The university is one of the fastest growing European universities in terms of research capacity and productivity, and its commitment to European research excellence is reflected by the recent extension of the 'European Office' in its Research Office, i.e. the department overseeing, guiding and administering research projects.

Curriculum vitae

Jeroen Demeyer (leadPI, male, 30 PM) is a post-doctoral assistant at Ghent University (Belgium). He has broad mathematical interests, mainly in number theory and its connections with logic. He received his PhD in mathematics in Ghent in 2007. Since then, he has been working as a post-doc in Ghent, except for a stay of 15 months at the Scuola Normale Superiore in Pisa (Italy). At Ghent University, he introduced SAGE in the mathematics education and teaches two courses with it.

He is a very active contributor to SAGE and he has been its release manager for a period of 3 years, from 2011 to 2013. He is the main author of the CYSIGNALS package for interrupt and signal support in CYTHON. He also made various contributions to other projects, such as PARI/GP, CYTHON, and JUPYTER.

Publications, products, achievements

At the Department of Mathematics of Ghent University, researchers have always been interested in computational aspects: Jeroen Demeyer has contributed to SAGE and Jan De Beule and Michel Lavrauw are co-authors of the GAP package FinInG for finite incidence geometry.

Significant infrastructure

Ghent University has a large High Performance Cluster infrastructure, consisting of 5 distinct clusters for various purposes, totalling 11328 CPU cores. On top of this, it also hosts the Tier-1 Flemish Supercomputer, a single cluster with 8448 cores, to be used for all research institutions in Flanders.

Ghent University is using SageMath for teaching in the mathematics education. To support this, there are two 12-core servers running a Sage Notebook server. There is also a VM running Jupyter Hub. Eventually, this should also be used for teaching purposes, but this is currently in an experimental phase.

4.1.17 XFEL: European XFEL GmbH (DE)

The European X-Ray Free-Electron Laser Facility GmbH is a limited liability company under German law. At present, 11 countries are supporting European XFEL through cash and in-kind contributions: Denmark, France, Germany, Hungary, Italy, Poland, Russia, Slovakia, Spain, Sweden, and Switzerland. The company is in charge of the construction and operation of the European XFEL, a 3.4 km long X-ray free-electron laser facility extending from Hamburg to the neighbouring town of Schenefeld in the German federal state of Schleswig-Holstein. Civil construction started in early 2009; the beginning of user operation is planned for 2017. With its repetition rate of 27,000 pulses per second and a peak brilliance a billion times higher than that of the best synchrotron X-ray radiation sources, it is expected that the European XFEL will enable the investigation of still open scientific problems in a variety of disciplines (physics, structural biology, chemistry, planetary science, study of matter under extreme conditions and many others).

European XFEL, a landmark on the ESFRI Roadmap, is a single site X-ray research infrastructure. When operational, 3 beamlines and 6 experiments will be available for scientific users. The SASE1 beamline comprises the instruments Single Particles, clusters, and Biomolecules and Serial Femtosecond Crystallography (SPB/SFX) and Femtosecond X-ray Experiments (FXE), SASE 2 includes Materials Imaging and Dynamics (MID) and High Energy Density Science (HED) and SASE3 Small Quantum Systems (SQS), and Spectroscopy & Coherent Scattering (SCS).

In the context of this proposal, Hans Fangohr has long standing experience in high performance computer simulation to advance science and engineering, and the education of researchers in the most effective pursuit of computational science.

European XFEL is using IPython and the Jupyter Notebook as core utilities in their large scale experiment control, data capture and data analysis.

Curriculum vitae

Hans Fangohr (leadPI, male, 3 PM) Hans Fangohr is Professor of Computational Modelling at the University of Southampton until August 2017 (3PM), then Senior Data Analysis Scientist at the European XFEL GmbH (3PM). 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`) at Southampton, 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.

Marijan Beg (R, male, 24 PM) Dr Marijan Beg is a post-doctoral researcher with experience in computational science, IPython and the Jupyter Notebook and micromagnetic simulations. He is working under the leadership of and together with Hans Fangohr at Southampton until August 2017 (16PM), and then moving to continue the work at European XFEL GmbH from September 2017 onwards (24PM).

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)

4.1.18 FAU: FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN/NÜRNBERG (DE)

Friedrich Alexander Universität Erlangen/Nürnberg (FAU) is a public research university in the cities of Erlangen and Nuremberg, Germany. FAU is the second largest state university in the state Bavaria. It has 5 faculties, 23 departments/schools, 30 clinical departments, 19 autonomous departments, 656 professors, and ca 40 000 students.

Prof. Dr. Michael Kohlhase has moved to FAU from JacobsUni 1. September 2016 and continues his research there.

The KWARC (KnoWledge Adaptation and Reasoning for Content [KWARC]) 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. The KWARC group moved to FAU over fall 2016. Its group working on OpenDreamKit will be composed of the following non-exhaustive list: Michael Kohlhase, Florian Rabe, Mihnea Iancu and Tom Wiesing. The total number of Person-Months will go up to 43 Person-Months according to the Table 3.4.1.

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).

Christian Maeder (Res, male, 19 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, 18 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 [Koh06] and redeveloped its formal core in the OMDoc/MMT format [RK13]. The latter has been implemented in the MMT system [MMT; RK13] 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[KŞ06; KMP12].

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 [JK10] as well as automatic annotation using linguistic techniques [Gin+09]. 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 [Koh+11; MH; Ian+14].

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.

JacobsUni 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.

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

CNRS

The University of Bordeaux will be involved as a Third Party Linked to CNRS. Indeed, both the LaBRI and the IMB are Joint Research Units for which both the University of Bordeaux and CNRS are supervisory authorities ; Karim Belabas and Adrien Boussicault are staff of the University of Bordeaux and actively participate in the project (see 4.1.2). The University of Bordeaux was founded in the 15th century and nowadays the city hosts two universities, dozens of schools and 100 research laboratories with partner institutions such as CNRS, Inserm, INRA and INRIA. It is an important centre of studies and research in France with approximately 50,000 students, 2,000 PhD students and 5,000 researchers.

Does the participant plan to subcontract certain tasks	No
Does the participant envisage that part of its work is performed by linked third parties	Yes
Two participants of the project are staff of the University of Bordeaux: Karim Belabas and Adrien Boussicault (see above).	
Does the participant envisage the use of contributions in kind provided by third parties	No

University of Silesia

Does the participant plan to subcontract certain tasks	Yes
See 3.4.2 for justification	
Does the participant envisage that part of its work is performed by linked third parties	No
Does the participant envisage the use of contributions in kind provided by third parties	No

Logilab

Does the participant plan to subcontract certain tasks	Yes
See 3.4.2 for justification	
Does the participant envisage that part of its work is performed by linked third parties	No
Does the participant envisage the use of contributions in kind provided by third parties	No

Other participants

There is no subcontracting costs for other participants.

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

[Bra+15] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, eds. *Handbook of Computational Social Choice*. Cambridge University Press, 2015.

[com08] T. Sage-Combinat community. *Sage-Combinat: enhancing Sage as a toolbox for computer exploration in algebraic combinatorics*. `http://combinat.sagemath.org`. 2008.

[Doc] *Docker – An open platform for distributed applications for developers and sysadmin*. http://docker.com. 2015.

[DPV14] 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.

[Far07] 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`.

[FH13] J. Fromentin and F. Hivert. "Exploring the tree of numerical semigroups". In: (2013). Hal: `00823339v2`.

[Fis+07] 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.

[Gin+09] 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/lamapun/pubs/AST09_LaMaPUn+appendix.pdf`.

[Ian+14] M. Iancu, C. Jucovschi, M. Kohlhase, and T. Wiesing. "System Description: MathHub.info". In: *Intelligent Computer Mathematics 2014*. Ed. by S. Watt, J. Davenport, A. Sexton, P. Sojka, and J. Urban. LNCS 8543. Springer, 2014, pp. 431–434. ISBN: 978-3-319-08433-6. URL: `http://kwarc.info/kohlhase/papers/cicm14-mathhub.pdf`.

[Ipya] *Interactive Python (IPython)*. http://ipython.org. 2014.

[Ipyb] *Notebook based interactive GUI for common simulation requirements*. http://www.youtube.com/watch?v=RhMHgQbP_A. 2014.

[JK10] C. Jucovschi and M. Kohlhase. "sTeXIDE: An Integrated Development Environment for sTeX Collections". In: *Intelligent Computer Mathematics*. Ed. by S. Autexier, J. Calmet, D. Delahaye, P. D. F. Ion, L. Rideau, R. Rioboo, and A. P. Sexton. LNAI 6167. Springer Verlag, 2010, pp. 336–344. ISBN: 3642141277. arXiv: `1005.5489v1 [cs.OH]`.

[Jup] *Project Jupyter*. http://jupyter.org. 2014.

[Ket] K. Ketcheson.

[KI12] M. Kohlhase and M. Iancu. "Searching the Space of Mathematical Knowledge". In: *DML and MIR 2012*. Ed. by P. Sojka and M. Kohlhase. Masaryk University, Brno, 2012. ISBN: 978-80-210-5542-1. URL: `http://kwarc.info/kohlhase/papers/mir12.pdf`.

[KMP12] M. Kohlhase, B. A. Matican, and C. C. Prodescu. "MathWebSearch 0.5 – Scaling an Open Formula Search Engine". In: *Intelligent Computer Mathematics*. Ed. by J. Jeuring, J. A. Campbell, J. Carette, G. Dos Reis, P. Sojka, M. Wenzel, and V. Sorge. 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`.

[KMR13] M. Kohlhase, F. Mance, and F. Rabe. "A Universal Machine for Biform Theory Graphs". In: *Intelligent Computer Mathematics*. 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`.

[Koh+11] 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`.

[Koh06] 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`.

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

[KŞ06] 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`.

[Mat] *MathOverflow*. URL: `http://mathoverflow.net` (visited on 09/08/2011).

[MH] *MathHub.info: Active Mathematics*. URL: `http://mathhub.info` (visited on 01/28/2014).

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

[MWS] *MathWebSearch – Searching Mathematics on the Web*. URL: http://search.mathweb.org (visited on 06/11/2016).

[Nis07] N. Nisan. “Introduction to mechanism design (for computer scientists)”. In: *Algorithmic game theory*. Cambridge Univ. Press, Cambridge, 2007, pp. 209–241.

[Nmaa] *Nmag — Guided tour*. http://nmag.soton.ac.uk/nmag/0.2/manual/html/guided_tour.html.

[Nmab] *Nmag micromagnetic simulation suite*. http://nmag.soton.ac.uk.

[Oom] http://math.nist.gov/oommf/.

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

[Reh10] J. Rehmeyer. *Massively Collaborative Mathematics*. Apr. 1, 2010. URL: http://www.siam.org/news/news.php?id=1731 (visited on 01/08/2015).

[RK13] 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.

[She] *CVE-2014-6271*.

[Sof13a] Software.intel.com. *Intel® Cilk™ Homepage*. https://www.cilkplus.org/. 2013.

[Sof13b] Software.intel.com. *Intel® Cilk™ Plus Reference Guide*. https://software.intel.com/en-us/node/522579. 2013.

[SV14] 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.

[tea] I. team.

[Traa] .

[Trab] *Trac Open Source Project*.

[Vir] *VirtualBox*. http://virtualbox.org. 2015.