



Weierstrass Institute for
Applied Analysis and Stochastics



How WIAS handles mathematical research data

Thomas Koprucki, Karsten Tabelow, and Ilka Kleinod

Situation of research data today

- tremendous growth in data generation (huge to massive data)
- modeling and simulation tools readymade to users from applied disciplines
- multi-dimensional, multi-resolution, multi-modal data
- from experiment, simulation or data analysis
- large, unstructured long tail
- plethora of specific data formats and technical standards

Drawbacks:

- letterpress-oriented publication culture (graphs not data)
- paper-oriented recognition of achievements

Emerging:

- Data publications
- Data repositories
- Data sharing culture

Data journals (launched 2-3 years ago)

The screenshot shows the Elsevier website for the Geoscience Data Journal. The browser address bar displays 'journals.elsevier.com'. The page features the Elsevier logo on the left and a navigation menu with links to 'Journals - Forschun...', 'Journal - Elsevier', 'ScienceDirect.com', 'Geoscience Data Jo...', 'Home : Scientific Data', and 'Home : Scientific Data'. The main content area includes a 'JOURNAL TOOLS' section with links for 'Get New Content Alerts', 'Get RSS feed', and 'Save to My Profile'. A 'JOURNAL MENU' section lists 'Journal Home', 'FIND ISSUES' (Current Issue, All Issues), and 'FIND ARTICLES' (Early View). The journal title 'Geoscience Data Journal' is prominently displayed in green, with the Wiley logo to its right. Below the title, it states '© John Wiley & Sons Ltd' and 'Each article is made available under the terms of the Creative Commons Attribution License'. A small image of the journal cover is shown, along with the text 'Edited By: Dr Rob Allan, Met Office, UK' and 'Online ISSN: 2049-6060'.

ELSEVIER

Home > Books & Journals > Data in Brief

Data in Brief

Editor-in-Chief: Hao-Ran Wang
[View full editorial board](#)

JOURNAL TOOLS

- [Get New Content Alerts](#)
- [Get RSS feed](#)
- [Save to My Profile](#)

JOURNAL MENU

- [Journal Home](#)
- FIND ISSUES**
 - [Current Issue](#)
 - [All Issues](#)
- FIND ARTICLES**
 - [Early View](#)

Geoscience Data Journal

© John Wiley & Sons Ltd

Each article is made available under the terms of the [Creative Commons Attribution License](#)

Edited By: Dr Rob Allan, Met Office, UK
Online ISSN: 2049-6060

Wiley

The screenshot shows the Scientific Data journal website. The header features the title 'SCIENTIFIC DATA' in large white letters on a dark blue background, with a search bar and 'Go' button on the right. Below the header is a navigation menu with links to 'Home', 'Archive', 'About', 'For Authors', 'For Referees', 'Data Policies', and 'Collections'. The main content area is titled 'Featured Data Descriptor' and features a large image of a forest. To the right of the image is a text block describing the featured data descriptor: 'Plant traits, productivity, biomass and soil properties from forest sites in the Pacific Northwest, 1999-2014' by Berner & Law, dated 19th January 2016. The text describes the use of computational models to explore climate change impacts. On the right side of the page, there is a section titled 'About Scientific Data' which describes the journal as an open-access, peer-reviewed journal for descriptions of scientifically valuable datasets. It mentions the primary article-type, the 'Data Descriptor', and provides links to 'Browse publications by date' and 'Browse publications by subject'.

SCIENTIFIC DATA

Search

[Advanced search](#)

[Home](#) | [Archive](#) | [About](#) | [For Authors](#) | [For Referees](#) | [Data Policies](#) | [Collections](#)

Featured Data Descriptor



Plant traits, productivity, biomass and soil properties from forest sites in the Pacific Northwest, 1999-2014

Berner & Law | 19th January 2016

Computational models are widely used to explore the impacts of climate change, and to assess potential responses. But, such models often rely

About *Scientific Data*



Scientific Data is an open-access, peer-reviewed journal for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

[Browse publications by date](#) ▶

[Browse publications by subject](#) ▶

Nature Publishing Group

Research Data Repositories for MMS



INFORMATION SERVICES


Contact us

DSpace Startseite / College of Science & Engineering / School of Mathematics

School of Mathematics

Volltextsuche:

Los



Research in the [School of Mathematics](#) is undertaken in a wide variety of areas of the mathematical sciences, including pure, applied, statistics, operational research as well as mathematical physics. The research environment is enhanced by the International Centre for Mathematical Sciences and by close collaboration with Heriot-Watt University through the Maxwell Institute.

Research groups include:

Suche

☒ DSpace Suche
☐ In diesem Bereich

MEIN BENUTZERKONTO

Einloggen

Registrieren

STÖBERN

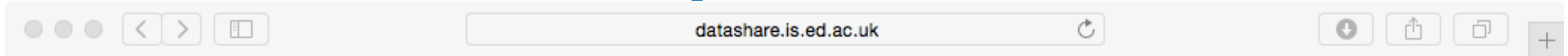
Gesamter Bestand

Bereiche & Sammlungen

Diesen Bereich

Titeln

Research Data Repositories for MMS



Neueste Zugänge

Eddy diffusivity diagnostics data

Mak, Julian; Maddison, James; Marshall, David

Constrained interpolation data

Maddison, James R.; Hiester, Hannah R.

Quasi-geostrophic double gyre force function data

Maddison, James R.; Marshall, D.P.; Shipton, J.

Parallel simplex benchmarking results

Huangfu, Qi; Hall, J.A.J.

Data used to generate results in "Parallelizing the dual revised simplex method"
<http://www.maths.ed.ac.uk/hall/HuHa13/>, a copy of which is attached.

Snapshots of an N-body model of M4

Heggie, Douglas (University of Edinburgh. School of Mathematics, 2014-09-04)

326 snapshots of an N-body model of the Galactic globular cluster M4. Each snapshot consists of a list of all the particles in the simulation at one time.

HypFun

Pearson, John (University of Edinburgh. School of Mathematics, 2014-07-29)

MATLAB code for computing the confluent and Gauss hypergeometric functions using a range of numerical methods.

Symmetry operators Mathematica notebook

Backdahl, Thomas (2014-04-17)

This is a collection of files intended for verification of the results in the paper Andersson, L, Bäckdahl, T, Blue, P,

Subject Classification

Astrophysics (1)

Mathematics (1)

Numerical Analysis (1)

... mehr

Subject Keywords

Numerical analysis, Special functions, Confluent hypergeometric function, Gauss hypergeometric function (1)

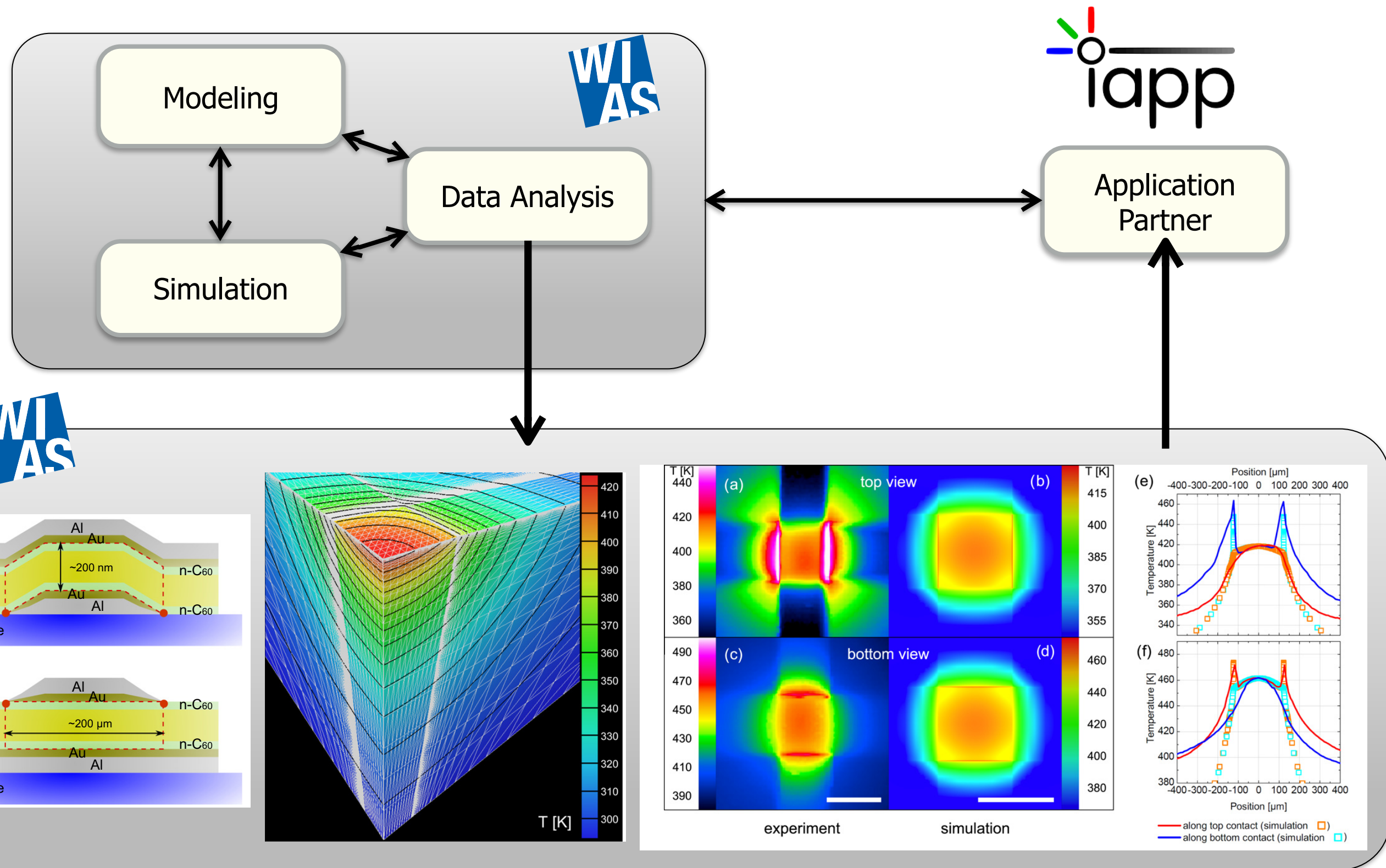
Revised simplex method, simplex parallelization (1)

Funder

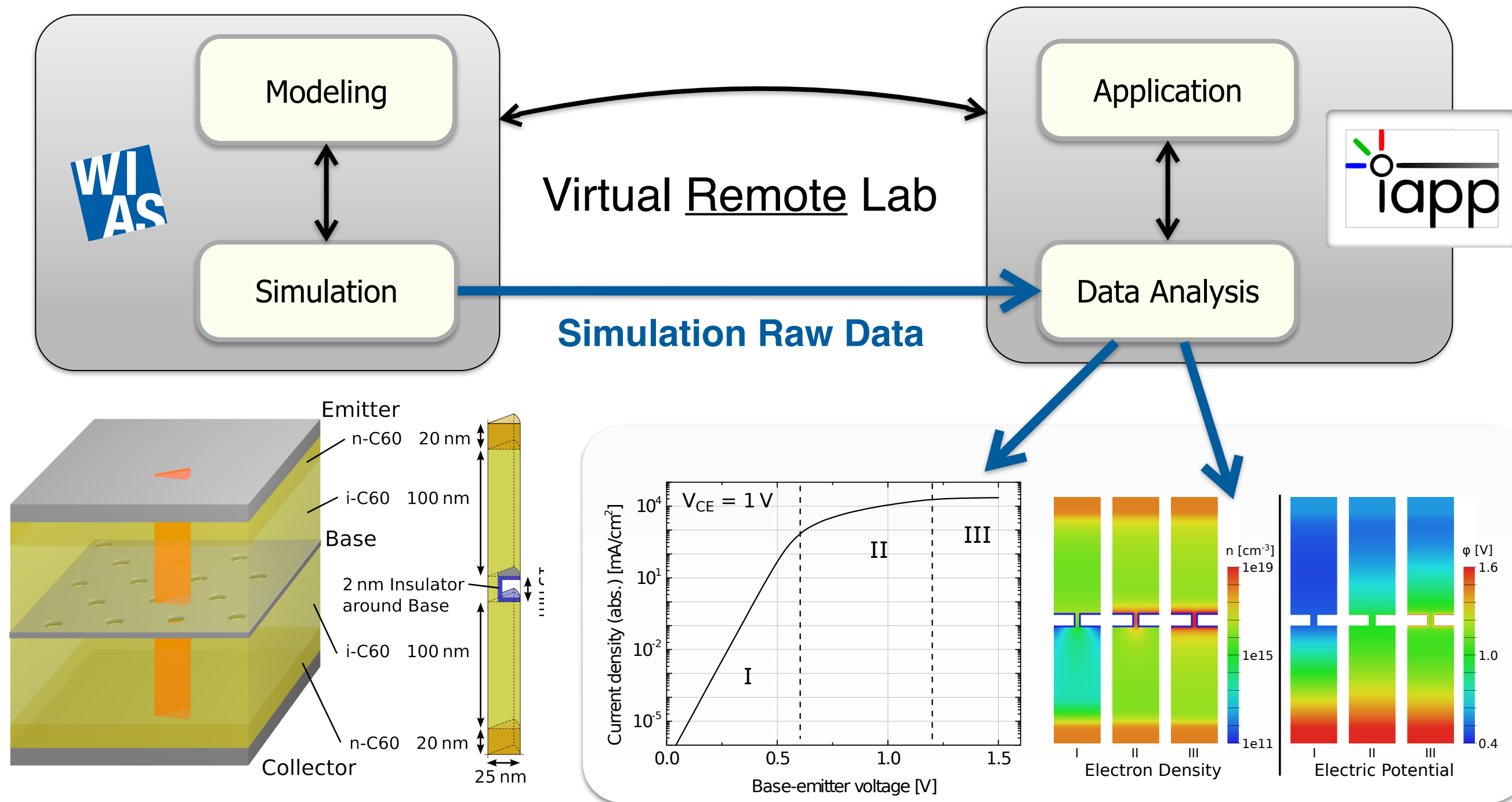
EPSRC - Engineering and Physical Sciences Research Council (4)

NERC - Natural Environment

Usual way of data sharing

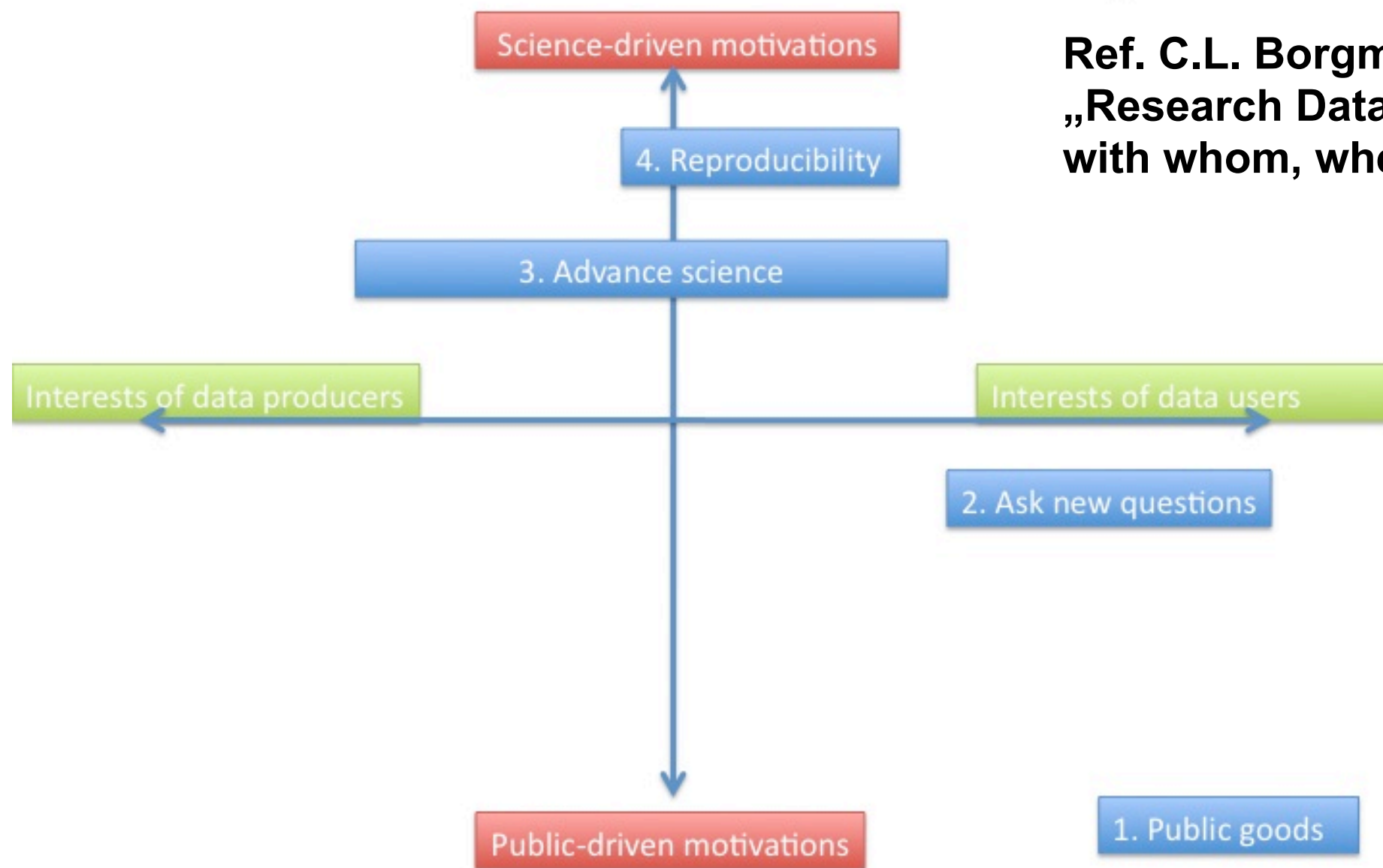


Added value of data sharing



better science by full exploitation of data

Motivation: Data Sharing

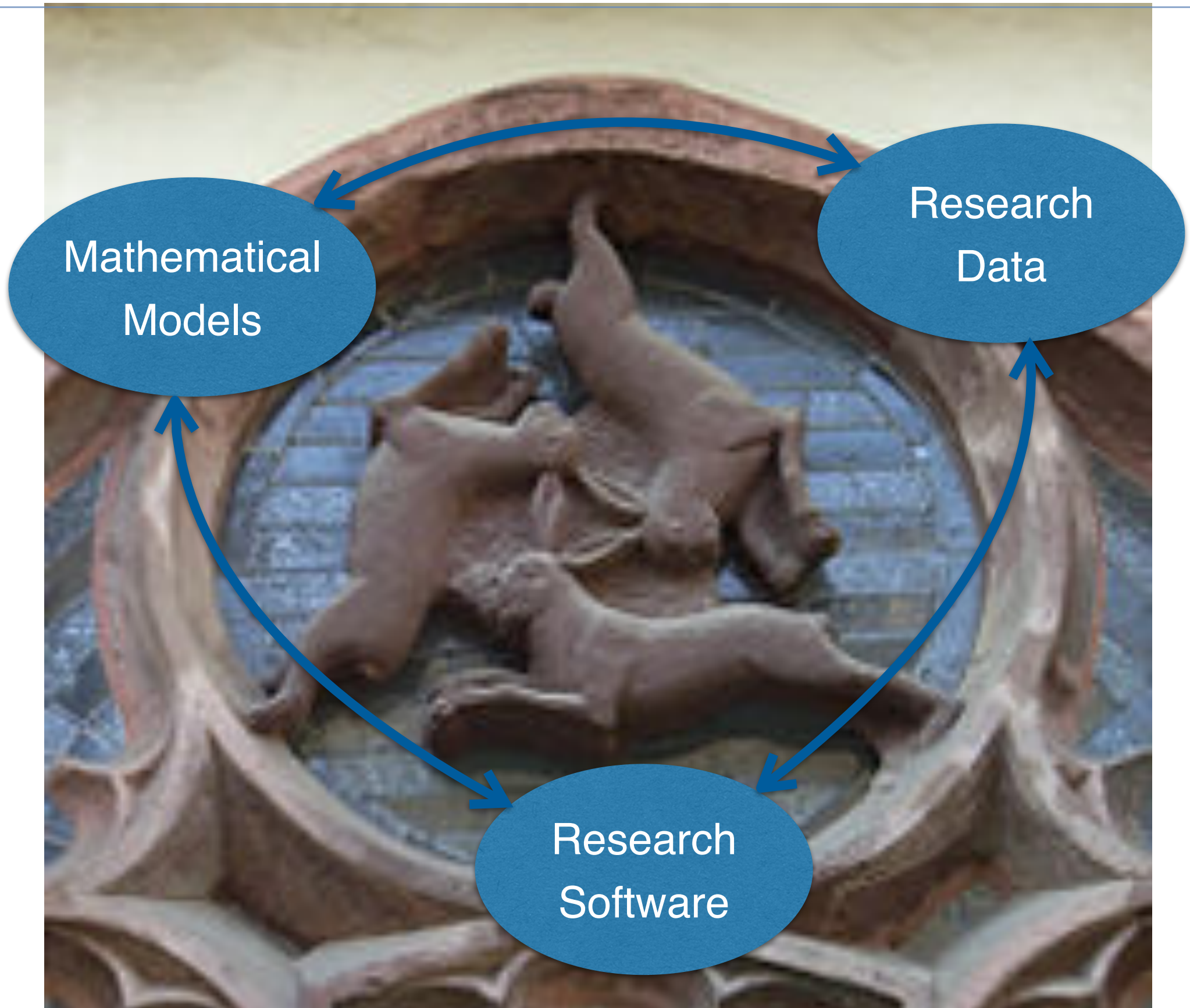


Ref. C.L. Borgmann (UCLA)
„Research Data: Who will share what,
with whom, when, and why?“, 2010

EU Commission:

[..] optimise the circulation, access to and transfer of scientific knowledge. [..]

Three rabbits



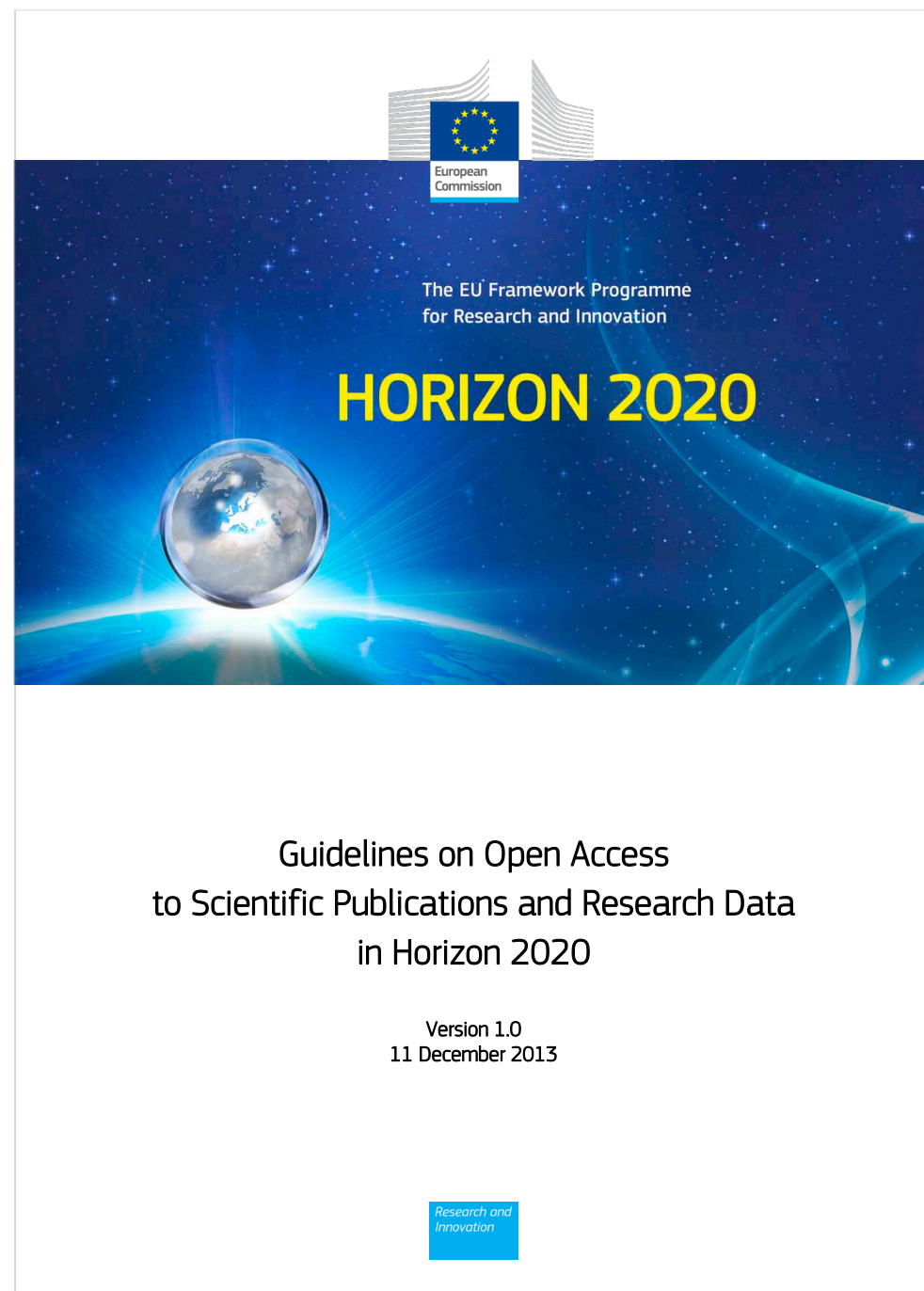
The ideal world



- Unique Identifier (DOI) for Models, Data and Software
- Smart Data Repositories (like Dropbox, ownCloud)
- Community specific but widely accepted standards and tools
- Digital Mathematical Models and Algorithms; Repositories
- Data-aware publication and recognition culture
- ePubs and Smart Services for interrelation and reuse

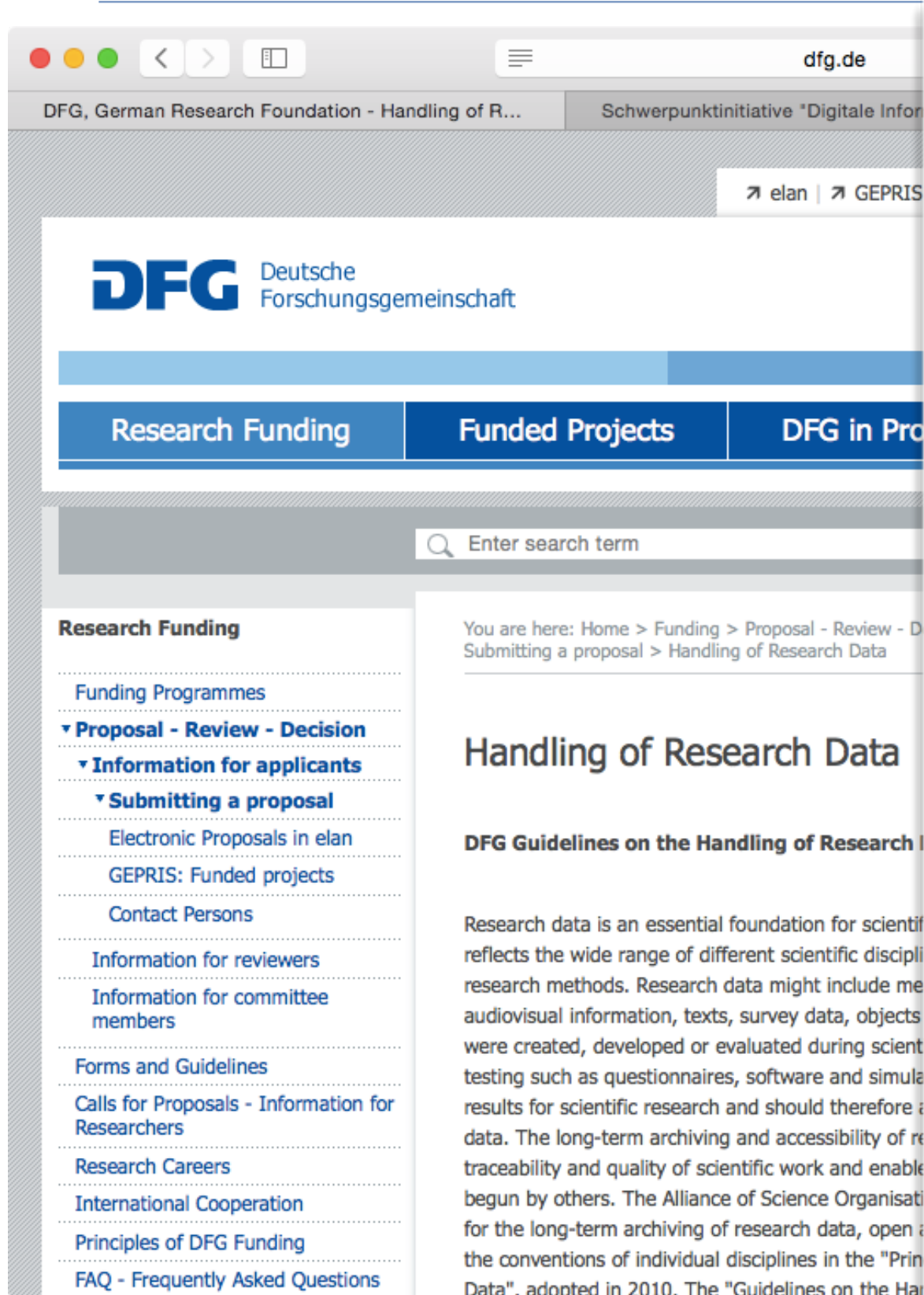
better data, better research

Institutions and Funding Organizations



OpenData Pilot in H2020 (for 20% of projects)
incl. Data Management Plan (DMP, mandatory) etc.

Institutions and Funding Organizations



Deutsche
Forschungsgemeinschaft

DFG Guidelines on the Handling of Research Data

Research data is an essential foundation for scientific work. The diversity of this data reflects the wide range of different scientific disciplines, research interests and research methods. Research data might include measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections, or samples that were created, developed or evaluated during scientific work. Methodical forms of testing such as questionnaires, software and simulations may also produce important results for scientific research and should therefore also be categorised as research data. The long-term archiving and accessibility of research data contributes to the traceability and quality of scientific work and enables researchers to carry on work begun by others. The Alliance of Science Organisations in Germany voiced its support for the long-term archiving of research data, open access to it and compliance with the conventions of individual disciplines in the "Principles for the Handling of Research Data", adopted in 2010¹. The "Guidelines on the Handling of Research Data" put the framework stipulated by the Principles into a concrete form in the DFG's funding arrangements.

The following general guidelines apply for applicants submitting proposals to the DFG:

1. Project planning and submission of proposals

Applicants should consider during the planning stage whether and how much of the research data resulting from a project could be relevant for other research contexts and how this data can be made available to other researchers for reuse. Applicants should therefore detail in the proposal what research data will be generated or evaluated during a scientific research project. Concepts and considerations appropriate to the specific discipline for quality assurance and the handling and long-term archiving of research data should be taken as a basis. The relevant explanations must contain information about data types, discipline-specific standards (if applicable) and the choice of suitable repositories, if these are available for a given research area or particular data types. Details should also be provided on any third-party rights affected and preliminary planning for the data publication schedule.

2. Accessibility

Assuming that the publication of research data from a DFG-funded project does not conflict with the rights of third parties (in particular data protection or copyright), research data should be made available as soon as possible. Data should be made accessible at a stage of processing that allows it to be usefully reused by third parties (raw data or structured data). To make sure this is the case, it must be ensured that access to the data is still guaranteed when, through publication, the rights of use relating to research data are transferred to a third party, usually a publishing house.

3. Long-term archiving

In accordance with the rules of good scientific practice, research data should be archived in the researcher's own institution or an appropriate nationwide infrastructure for at least 10 years.

The DFG offers the following assistance for the implementation of the guidelines:

1. Support and advice

The reuse of research data is playing an increasingly important role in nearly all scientific disciplines. At the same time, there are considerable differences between disciplines with regard to the current state of discussion and the available infrastructures. To help with the planning of research projects, information and suggestions for the handling of research data have been compiled here:

http://www.dfg.de/en/research_funding/proposal_review_decision/applicants/submitting_proposal/research_data/i

Institutions and Funding Organizations

From preamble of “DFG Guidelines on Handling of Research Data” (Adopted by the Senate of the DFG at Sep. 30, 2015):

- “Research data is an essential foundation for scientific work.”
- “The diversity of this data reflects the wide range of different scientific disciplines, research interests and research methods.”
- “Research data might include
... audiovisual information, texts, ...”
- **“... software and simulations may also produce important results for scientific research and should therefore also be categorised as research data.”**
- “The long-term archiving and accessibility of research data contributes to the traceability and quality of scientific work and enables researchers to carry on work begun by others.”

DFG Guidelines on Handling of Research Data

Research data is an essential foundation for a wide range of different scientific disciplines, research interests and research methods. It includes measurement data, laboratory values, collections, or samples that were created, developed, or tested, such as questionnaires, software, or simulations. Research data should therefore also be made accessible and its long-term archiving should be supported. The accessibility of research data contributes to the traceability and quality of scientific work and enables researchers to carry on work begun by others. The DFG's funding arrangements for research data should therefore be based on the conventions of individual disciplines in the field. The "Guidelines on the Handling of Research Data" take concrete form in the DFG's funding arrangements.

The following general guidelines apply for applicants:

1. Project planning and submission of proposals
Applicants should consider during the planning phase whether data from the project could be relevant for other research projects for reuse. Applicants should therefore detail in the proposal the handling of research data during a scientific research project. Concepts and plans for the handling of research data, assurance and the handling and long-term archiving of research data, and explanations must contain information about data repositories, if these are available for reuse, or suitable repositories, if these are available for reuse, or suitable repositories, if these are available for reuse, or suitable repositories, if these are available for reuse.

2. Accessibility
Assuming that the publication of research data is in the public interest, applicants should make data accessible (in particular data protection or copyright law). Data should be made accessible at a stage of the project when it is still possible to do so (e.g. data or structured data). To make sure this is the case, applicants should, when, through publication, the rights of use are transferred to the publishing house.

3. Long-term archiving
In accordance with the rules of good scientific practice, research data should be archived in a suitable institution or an appropriate nationwide infrastructure.

The DFG offers the following assistance for the handling of research data:

1. Support and advice
The reuse of research data is playing an increasingly important role. At the same time, there are considerable differences between the handling of research data in different disciplines. To help with the planning and handling of research data, the DFG has compiled a directory of research data repositories, a summary of other, quality-assured infrastructures, and a summary of other, quality-assured infrastructures. A directory of research data repositories is available at http://www.dfg.de/en/research_funding/proposals_and_projects/research_data/index.html. A summary of other, quality-assured infrastructures is available at <http://risources.dfg.de/>.

2. Costs for data preparation and for the use of research data
Applicants may request funding for project-specific costs for the preparation of research data for subsequent reuse. This funding can be requested as part of a proposal to the DFG. It is also possible to request funding for the use of research data in other projects. Financial support is available for the use of research data in other projects.

Institutions and Funding Organizations

Abgeordnetenhaus B E R L I N

17. Wahlperiode

Drucksache 17/2512

21.10.2015

Mitteilung – zur Kenntnisnahme –

**„Open-Access-Strategie für Berlin: wissenschaftliche Publikationen für jedermann
zugänglich und nutzbar machen“**

– Schlussbericht – Drucksachen 17/1487, 17/1655 und 17/2024

Institutions and Funding Organizations

From “Drucksache 17/2512, 21.10.2015 des Abgeordnetenhaus Berlin”

- “Forschungsdaten: **International weniger weit entwickelt ist das Handlungsfeld „Open Access zu Forschungsdaten“.**
Die Einrichtungen des Landes Berlin bilden hier keine Ausnahme.”
- “Der **offene Zugang zu Forschungsdaten** und deren **umfassende Nachnutzung** sollen gewährleistet werden.
Erforderliche Beiträge des Landes Berlin können nur im Rahmen von national und international abgestimmten Strategien geleistet werden, die noch formuliert werden müssen.”
- “Die Open-Access-Strategien der **wissenschaftlichen Einrichtungen in Berlin sollen eine Forschungsdaten-Policy enthalten**; die Wissenschaftlerinnen und Wissenschaftler sollen ausdrücklich ermutigt werden, die **Daten zitierfähig zur Verfügung zu stellen.** “
- “Für den so genannten *Long-Tail* der Forschungsdaten – also Datensätze mit geringem Volumen, die in **verschiedenen Datenformaten** vorliegen und **schwer standardisierbar** sind – **fehlen in vielen Fachgebieten angemessene Infrastrukturen.**”

Abgeordnetenhaus **BERLIN**

17. Wahlperiode

Mitteilung – zur Kenntnisnahme –

„Open-Access-Strategie für Berlin: wissenschaftlich
zugänglich und nutzbar machen“
– Schlussbericht – Drucksachen 17/1487, 17/1655 und

Steps at WIAS

- Research software and data (and mathematical models) play an important role for the transfer of our research to industry and other disciplines
- Workgroup to develop an Research Data Policy (since end 2014)
- Based on experiences of formulation of a software policy (adpoted at WIAS in May 2015)
- discussion on the topic at WIAS
 - definition for math (!)
 - best practices and guidelines
 - technical infrastructure