

**Please address the following aspects in your letter of intent**

## 1 Binding letter of intent as advance notification or non-binding letter of intent

*[Please indicate clearly whether your document is a binding letter of intent as advance notification or a non-binding letter of intent.]*

<input checked="" type="checkbox"/>	Binding letter of intent (required as advance notification for proposals in 2019)
<input type="checkbox"/>	Non-binding letter of intent (anticipated submission in 2020)
<input type="checkbox"/>	Non-binding letter of intent (anticipated submission in 2021)

## 2 Formal details

- **Planned name of the consortium**

National Research Data Infrastructure for Mobility Technology

- **Acronym of the planned consortium**

NFDI4MobilTech

- **Applicant institution**

Karlsruhe Institute of Technology (KIT), Kaiserstraße 12, 76131 Karlsruhe

Prof. Dr.-Ing. Holger Hanselka

- **Spokesperson**

Prof. Dr. rer. nat. Frank Gauterin, [frank.gauterin@kit.edu](mailto:frank.gauterin@kit.edu),

KIT, Institute of Vehicle System Technology (FAST)

NFDI4MobilTech was initiated by Prof. Frank Gauterin, KIT and Prof. Sahin Albayrak, TU Berlin. A dual chairmanship of KIT and TU Berlin is planned.

As requested, co-applicant institutions, co-spokespersons and participants will be submitted until August 16<sup>th</sup>, 2019.

## 3 Objectives, work programme and research environment

- **Research area of the proposed consortium (according to the DFG classification system):**

44: Computer Science, Systems and Electrical Engineering

## Concise summary of the planned consortium's main objectives and task areas

### Main Objectives

The NFDI 4 MobilTech initiative aims to create data management structures for the domain specific challenges of ground-based transport and its travel behaviour aspects with detailed consideration of technical systems. Vehicle generated data currently tend to be unavailable to research institutions – manufacturers are very reluctant to share it. Hence, a sound data repository for scientific use, combining data generated by research vehicles, test fields, mobile devices, etc. of research institutions in Germany and Europe is particularly valuable. The platform should allow for search and analysis at different levels of abstraction, from raw data to high-level descriptions of scenes and scenarios.

### Task Areas

Data Acquisition: Based on a survey of different data sources, formats, semantic contexts, and data acquisition processes in the community, a set of recommendations and tools is generated to do data acquisition in a way that allows an easy analysis across repositories. This addresses standards for data formats, naming, documentation, selection of data to be stored, etc. It includes automatic metadata generation and labelling tools.

Data Pre-processing: For a successful automated feature extraction, multi-layered mobility data need pre-processing and cleaning. Data from multi-sensor setups have to be aligned, synchronized, and linked. All privacy-related information has to be identified and either removed or anonymized. Reference methodologies such as KDD pipeline will be used.

Feature Extraction: Based on the abstracted input data, a framework is developed to extract mid- to high-level features. Based on operator-defined criteria, various sources or channels of input data can be processed to yield fused information at various levels of abstraction, from load profiles or time-to-collision (TTC) up to semantic quantifiers as accident risks, or tags such as “near-collision” for use in database queries.

Data Analysis: We consider data coming from a variety of domains and sources, such as autonomous cars, simulation environments, real-world test beds amongst others. The challenge we are addressing is to support users to determine the best suitable analysis method for a specific task in a given application domain and to integrate analysis methods and their results going beyond specific tasks across application domains.

Use and Reuse: For the platform, it is necessary to elicit and analyse stakeholders, usage scenarios, quality requirements, and constraints. Currently, there is a barrier between available data and information needs of potential users. Important aspects include data privacy, data quality, and accessibility for documented use, scientific evaluations, or use in teaching. Goal is the implementation of an operating model for the NFDI platform.

Archiving: For archiving of research data duration, format, storage media, degree of availability, location, access restrictions, and licenses must be defined. A regulated evaluation decides about short- or long-term archiving. For quality assurance, the completeness of data, metadata, and documentation must be ensured. Archiving must be sufficiently documented to ensure complete reproducibility and traceability. Hard- and software required for subsequent reuse of data should be archived as well.

Reference Architecture for MobilTech: We will design a software architecture that allows the administration of MobilTech data and foresees structured processes for its adaptation and extension and offers functionality to cover sophisticated information and analysis needs. We will prototypically implement selected parts of this architecture. It will serve as a master for a robust and sustainable implementation by a third-party service provider.

Metadata Schema: A data inventory will be created in order to evaluate the available and known data sources, standards, and data models used by the community and partners. Based on this evaluation, a metadata schema would be either defined or selected from available standards, paying special attention to project requirements and collaborations with other NFDI initiatives.

Consistency / Data Integration: To deal with the fact, that no unified data model exists (neither for autonomous vehicles, nor for mobility systems) we work on mechanisms, tools, and processes to allow unified access to such coupled data models.

Access Control and Authorization: We will develop concepts to ensure that only authorized users can access sensitive data. To provide security even if a database server is hacked, data access should only be possible when an authorized user and the database server work together.

User Involvement and User Orientation: User involvement is a key to continuously adapt NFDI4MobilTech to changing requirements. Easy to use online feedback and voting tools, user community integration into decision making, conferences and workshops, adaptive user interfaces, user support and training will be implemented. Data providers will benefit from proper scientific credits and visible acknowledgements.

Networking: We aim to increase the visibility and acceptance of the platform and the general idea in the target community. We will involve the community and its institutions by a broad communication approach. We will act as a sounding board to economy and politics. The national and international partner base will be extended continuously.

Governance: A participative governance structure will be implemented, making use of the above mentioned instruments for user involvement. The (co-)spokespersons build a management team with dual chairmanship which reports to the partner assembly in regular user conferences and will be supported by an advisory board.

### **Brief description of the proposed use of existing infrastructures, tools and services**

**Infrastructures:** Sources which will provide research data for scientific use by the community: *Test Areas*: Digital Test Field A9 (BUM<sup>1</sup>), autonomous drive and traffic control track, Munich (BUM<sup>1</sup>), Application Platform for Intelligent Mobility – AIM (DLR), Test Bed Lower Saxony (DLR), smartphone mobility data MovingLab (DLR), Test Area Autonomous Driving Baden-Württemberg - TAF (FZI<sup>1</sup>, KIT), Karlsruhe Applied Machine Learning Living Lab – KarAMeL (FZI<sup>1</sup>), Living Lab Automotive & Living Lab smartMobility (FZI<sup>1</sup>), digital test field SAFARI, Berlin (FU Berlin), test road infrastructure for autonomous driving DIGINET-PS (TU Berlin), automated city driving environment “StadtPilot” (TUBS<sup>1</sup>), August Euler Air Field for automated vehicle and tram test (TUD<sup>1</sup>), Information and Communication Technology Infrastructure for Connected and Automated Road Transport - ICT4CART (UU<sup>1</sup>). *Test Stands*: X-in-the-loop test stand network XiL BWe (KIT, UU<sup>1</sup>), vehicle test stands (KIT, TU Berlin). *Automated vehicles*: (DFKI<sup>1</sup>, DLR, IOSB<sup>1</sup>, FU Berlin, FZI<sup>1</sup>, KIT, TU Berlin, TUBS<sup>1</sup>, TUD<sup>1</sup>, UU<sup>1</sup>). *Repositories*: Clearing House for Transport Data (DLR), Data sets (DLR), German Mobility Panel MOP (KIT), KITTI Vision Benchmark Suite (KIT), TUBS Road User Dataset (TUBS<sup>1</sup>), DriveU Traffic Light Dataset DTLD (UU<sup>1</sup>). Computing resources which will be used for central NFDI user access, data management services, and interlink of distributed repositories of the partners: AIM Backend (DLR), Repository for archival and publication of research data, data privacy and anonymization RADAR (FIZ<sup>1</sup>), FORDATIS (Fraunhofer), Large Scale Data Facility - LSDF: multi petabyte storage for research data (KIT), PetaFlop/s Supercomputing - ForHLR, bwUniCluster (KIT), long time storage of research data bwDataArchive Service (KIT).

**Tools** which are open to the community or will be used as blueprint for NFDI tools:

Helmholtz Analytics Toolkit – HeAT (DLR, KIT), sensor data tool FROST (IOSB<sup>1</sup>), video and lidar data labelling and visualisation tool (KIT), labeling tool for object annotation in laser scans and videos (TUBS<sup>1</sup>), full software stack for autonomous driving (FU Berlin), data management plan creation tool RDMP (TUBS<sup>1</sup>), Mobile Edge Computing based object recognition - MEC-View (UU<sup>1</sup>).

**Services** which are open to the community or will be used as blueprint for NFDI services:

Smart Data Innovation Lab provides hard- and software for the whole data lifecycle – SDIL (KIT), virtual machines - bwCloud SCOPE (KIT), research data management services - RDM@KIT (KIT), TUBS.researchdata competence center offers data consultation, training and project support, and training material (TUBS<sup>1</sup>).

<sup>1</sup> BUM Bundeswehr University Munich, DFKI German Research Center for Artificial Intelligence, FIZ Leibniz-Institut für Informationsinfrastruktur, FZI Forschungszentrum Informatik Karlsruhe, IOSB Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, TUBS TU Braunschweig, TUD TU Darmstadt, UU University of Ulm

**Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration****Agreement with NFDI4Ing**

The agreement for shared tasks with NFDI4Mobiltech includes annual workshops, joint working groups and coordination of community activities in the subject area transport and mobility, with the DLR acting as a joint between the two consortia to ensure a reliable flow of information. The focus of NFDI4Mobiltech will be on fine-grained, decidedly domain specific challenges of ground-based transport and its travel behaviour aspects, while NFDI4Ing will enable the integration of engineering-specific requirements in the mobility domain, including air-borne transport, with other engineering disciplines based on their common denominators.

**Planned collaboration with other NFDI initiatives**

- ForumX: We intend to create synergies to the planned repository “x-science” for data from experimental social sciences.
- NFDI4Earth: Data from NFDI4MobilTech such as road topography or rain sensors can be used for earth model validation.
- OD-Rex: Machine vision and learning, autonomous decision making, trajectory planning, online parameter estimation, real time capable optimization methods, etc. are of high importance for both, autonomous mobility systems and robotics. We expect fruitful collaboration concerning standards for metadata formats and many more.

**4 Cross-cutting topics**

**Please identify cross-cutting topics that are relevant for your consortium and that need to be designed and developed by several or all NFDI consortia.**

**Reference Architecture for MobilTech**

A platform architecture to operate central tasks and to interlink local repositories of the partners will be necessary for the NFDI4MobilTech consortia. Other consortia may have similar needs which would be a basis for collaboration.

### Anonymisation, Data Privacy, Copyright

The anonymization of data with critical content and personal information is not limited to the area of mobility: The anonymization of image data, but also the associated additional metadata, is also highly relevant in medical applications.

### Access Control and Authorization

For every project that relies on storing sensitive data, access control and authorization are critical topics that must be addressed. Since access control abstracts from the content of the data to be protected, it seems sensible to standardize measures across different initiatives.

### Metadata Concepts

A basic set of formatted metadata from which appropriate metadata can be selected according to the needs of the community and the specific use case could be very supportive for synergetic cross-domain data usage.

Important interdisciplinary challenges in collecting and handling research data are the consistent representation of uncertainties and correlations between measurements, and a means of expressly incorporating, but clearly distinguishing between actual measurements, processed/fused values, estimates, assumptions, simulated data and simulation parameters. A sound and transparent modeling of the actual and stochastic information content of the data can enable their use in validation of safety-critical technologies, such as medicine, civil engineering, cybersecurity, and mobility.

**Please indicate which of these cross-cutting topics your consortium could contribute to and how.**

### Reference Architecture for MobilTech

We seek a design where domain-specific aspects are not hard-coded, but – for instance in the form of process descriptions – easily exchangeable. So this reference architecture may serve at least partially as a blueprint for other scientific communities with similar boundary conditions such as many distributed data providers and an even wider group of users or it may be used as a master for a robust implementation by a third-party service provider.

### Anonymisation

Anonymisation of mobility data does not just mean to hide licence plates or vehicle labeling in video records but to avoid conclusions on the personality of travellers based on repeating patterns of movement or specific combination of attributes. An automated search for such indications requires more sophisticated tools which might also be relevant for other consortia.

### Metadata Concept

After the development, NFDI4MobilTech can provide its own proven metadata schema to the community of NFDI consortia in order to identify a basic set of standardised metadata, which can be selected appropriately to the use case.