Large-scale EXecution for Industry & Society

Deliverable D9.2

IPR and Data Management Approach

Co-funded by the Horizon 2020 Framework Programme of the European
Grant Agreement Number 825532
ICT-11-2018-2019 (IA - Innovation Action)

DELIVERABLE ID | TITLE	D9.2 | IPR and Data Management Approach
RESPONSIBLE AUTHOR	Florin Appopei (TESEO)
WORKPACKAGE ID | TITLE	WP9 | Impact on targeted sectors
WORKPACKAGE LEADER	TESEO
DATE OF DELIVERY (CONTRACTUAL)	30/06/2019 (M06)
DATE OF DELIVERY (SUBMITTED)	01/07/2019 (M07)
VERSION | STATUS	V1.1 | Final
TYPE OF DELIVERABLE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	All partners
INTERNAL REVIEW	Donato Magarielli (Avio Aero), James Hawkes (ECMWF)

Project Coordinator: Dr. Jan Martinovič – IT4Innovations, VSB – Technical University of Ostrava
E-mail: jan.martinovic@vsb.cz, **Phone:** +420 597 329 598, **Web:** https://lexis-project.eu

DOCUMENT VERSION

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Table of Content	21/05/2019	Florin Ionut Apopei (TESEO), Kateřina Slaninová (IT4I), Jan Martinovič (IT4I)
0.2	Pilots contribution	31/05/2019	Donato Magarielli (Avio Aero), Thierry Goubier (CEA), Antonio Parodi (CIMA)
0.3	Comments and changes of Section 2	17/06/2019	Vojtěch Muller (IT4I), Kateřina Slaninová (IT4I)
0.4	First review	18/06/2019	Donato Magarielli (Avio Aero)
0.5	LINKS contributed to Section 3.3	14/06/2019	Alberto Scionti (LINKS)
0.6	Review of Section 3.3 (proof. read, minor reformatting and updates)	19/06/2019	Marc Levrier (ATOS)
0.7	ECMWF review	21/06/2019	James Hawkes (ECMWF)
0.8	Industrial partners review, update of introduction to Section 3	27/06/2019	Sean Murphy (CYC), Stephan Hachinger (LRZ)
0.81	Approval of IPR section by all partners	27/06/2019	VPI group members
1.1	Final review by coordinator	30/06/2019	Jan Martinovič (IT4I), Kateřina Slaninová (IT4I)

GLOSSARY

ACRONYM	DESCRIPTION
AAI	Authorization and Authentication Infrastructure
BB	Burst Buffer
CA	Consortium Agreement
CAE	Computer Aided Engineering
CFD	Computational Fluid Dynamics
DMP	Data Management Plan
DSS	Data Science Storage
EAB	External Advisory Board
EU	European Commission
FAIR	Findable, Accessible, Interoperable and Re-usable
GA	Grant Agreement
HD	Hardware
HPC	High Performance Computing

HW	Hardware
IPR	Intellectual Property Rights
ISMS	Information Security Management System
LTDS	Let the Data Sing
MPCDF	Computing Centre of the Max Planck Society
RPO	Required Recovery Point Objective
RTO	Required Recovery Time Objective
SMS	Service Management System
SW	Software
WCDA	Weather and Climate Data API
WP	Work Package

TABLE OF PARTNERS

ACRONYM	PARTNER
Avio Aero	GE AVIO SRL
AWI	ALFRED WEGENER INSTITUT HELMHOLTZ ZENTRUM FUR POLAR UND MEERESFORSCHUNG
BLABS	BAYNCORE LABS LIMITED
Bull/Atos	BULL SAS
CEA	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES
CIMA	Centro Internazionale in Monitoraggio Ambientale - Fondazione CIMA
CYC	CYCLOPS LABS GMBH
ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
GFZ	HELMHOLTZ ZENTRUM POTSDAM DEUTSCHESGEOFORSCHUNGSZENTRUM GFZ
IT4I	VYSOKA SKOLA BANSKA - TECHNICKA UNIVERZITA OSTRAVA / IT4Innovations National Supercomputing Centre
ITHACA	ASSOCIAZIONE ITHACA
LINKS	FONDAZIONE LINKS / ISTITUTO SUPERIORE MARIO BOELLA ISMB
LRZ	BAYERISCHE AKADEMIE DER WISSENSCHAFTEN / Leibniz Rechenzentrum der BAdW
NUM	NUMTECH
O24	OUTPOST 24 FRANCE
TESEO	TESEO SPA TECNOLOGIE E SISTEMI ELETTRONICI ED OTTICI

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

As the first version, this deliverable aims to set up the initial strategy for the IPR and Data Management Plan, taking into account aspects related to the whole project, and clarifying the policies about data which will be part of the LEXIS platform and coming out of the 3 pilots (Aeronautics, Weather & Climate and Earthquake & Tsunami).

Position of the deliverable in the whole project context

The position of this document is the first of two deliverables focusing on the IPR and Data Management approach and its main goal is to set up the guidelines and strategy for handling the IPR and clarify the Data approach during the project. The second version will be delivered in M30 and will report all the results reached at the end of the project in term of IPR domain and Data coming out from pilots and the LEXIS platform.

Description of the deliverable

This deliverable focuses on the IPR strategy (Section 2) and Data Management Plan (DMP) (Section 3).

After the Introduction (Section 1), Section 2 describes the initial strategy of the project related to the IPR protection applicable to the LEXIS project in terms of exploitation and dissemination of results as well.

Section 3 describes DMP for the LEXIS project. The section covers data management policies of the supercomputing centres as main providers of LEXIS infrastructure, LEXIS platform data management overview including a workflow orchestration system, data management through the devised LEXIS data system, and front-end portal, and data management plans for the LEXIS pilots.

1 INTRODUCTION

Correctly handling the Intellectual Property Rights (IPR) in a Horizon 2020 project is crucial, especially if there are market-facing products arising from the project. For this reason, the European Commission (EU) provides guidelines for handling IPR from the proposal stage to the end of the project to manage the open source-software and non-open as well as will be entailed in Section 2.6.

Due to the importance of IPR, the LEXIS consortium has described how to manage IPR in the proposal and agreed the IPR management in the Consortium Agreement (CA) and the Grant Agreement (GA). The consortium has also decided upon the main roles concerning IPR management and followed the IPR Help-desk guidance rules for defining the IPR domain for the LEXIS project.

With this in mind, the aim of this deliverable is to set up the initial IPR strategy for LEXIS, taking into account all the IPR requirements that apply to the LEXIS project domain.

The person responsible for the management of this process is the Innovation Manager, who is also the task leader of IPR in T9.4. The other partners involved in managing IPR will be BLABS, as Task Leader of Exploitation, and IT4I as Project Coordinator.

Under this responsibility, there is also the process to manage third parties and let them use LEXIS developed software. To the third parties, the licences to such developer software will be provided. This process will be duly described in Section 2.6.

The tools available for IPR protection are listed in the IPR guidelines written by the European IPR HelpDesk and those relevant to LEXIS are described in Section 2 of this deliverable. If non-open-source tools or company tools are used to generate some results, these tools and their related licenses will be listed.

TESEO has previous experience and interaction with the IPR HelpDesk and this will be employed to implement the LEXIS IPR strategy, covering IPR management of the LEXIS platform and the individual pilots.

This process will be documented in the next update of this deliverable. D9.6 - Report on IPR Management in M30 [1] will report and update the evolution at that stage of the project related to IPR protection coming out from the project.

As agreed in the GA and the CA, all the mutual responsibilities and results will be duly treated with respect to the IPR policy.

There are a variety of IPR protection methods that could be employed in LEXIS to meet these requirements:

- Trademarks (a particular word or sign to distinguish a product or some products),
- Patents (exclusive rights granted for the protection of inventions – products or processes),
- Industrial Design (a right that protects the visual design of objects that are not purely utilitarian),
- Utility models (a right over the commercialisation of a protected invention),
- Trade secrets (confidential information can be protected by a trade secret),
- Copyright (the right of the creator over his invention),
- Domain names (is the right to protect the internet address and locate a particular “subject” in a particular uni-vocal way on the net),
- Geographical indications (a sign used on products having a specific geographical origin and whose qualities and/or reputation are attributable to that origin).

The specific methods which will be employed by LEXIS will be duly analysed and detailed in the following section.

To be able to manage in the best way the IPR in LEXIS, TESEO should organise a training session with the IPR HelpDesk in order to better perform in the management of the IPR and create an important know-how. All the involved partners in the IPR process will be included in. TESEO has already done such experience and the training session could be organised via webinar or physically, in both cases there will be present an expert provided by the

EU IPR Help-desk in order to better comprehend the whole project and the eventual issues to be managed in relation to IPR.

The final part of this document (Section 3) is focused on handling the data coming out from the project and, in particular, the strategy to make it compliant with IPR protection methods. Data Management Plan (DMP) describes the context in which research data will be generated and how they will be managed, maintained and preserved. DMP for the LEXIS project was prepared according to the Guidelines on FAIR Data Management in H2020 (Version 3 dated 26/06/2016)[1]. The section covers data management policies of the supercomputing centres as main providers of LEXIS infrastructure, LEXIS platform data management overview including a workflow orchestration system, data management through the devised LEXIS data system, and front-end portal, and data management plans for the LEXIS pilots.

2 PROJECT IPR STRATEGY

After the initial stage of the project and the signature of the CA and GA, it is the responsibility of LEXIS to implement an IPR strategy which meets the relevant requirements.

This strategy must tackle the following points:

- Access rights,
- Results and joint ownership,
- Transfer, protection, exploitation and dissemination of results,
- Maintaining confidentiality,
- Reporting,
- Post-project obligations.

The strategy aims to cover the whole project but could be divided into two different levels of applicability: at the level of the whole project and at the level of the individual pilots.

Another goal is to handle the data coming out from the pilots in the correct way, for this reason for example Avio Aero has previewed to use Dummy data to show up the functionality of the Aeronautics Pilot.

As agreed in the GA and CA, all the partners in the consortium must give access rights to their background and results created with other partners, in order to carry out their work on the project and exploit their results but always taking into account the specific limitations or conditions for implementation and exploitation as already described in the CA.

Special consideration must be applied to results born out of Joint Ownership (generated by at least two partners together). This specifically applies when the respective contribution of each beneficiary cannot be determined, or it is not possible to separate these contributions for the purposes of applying for, obtaining, or maintaining their protection.

Protection of results is crucial, but the protection depends on the nature of the result. Table 1 illustrates the type of result and the protection to be applied.

At this stage, the consortium aims to define what could be part of the IPR protection and explains the strategy to handle all aspects of the IP domain. The consortium has already established protection by Trade Secret (specifically a Non-Disclosure Agreement) for the partners involved in WP5 and will continue its role throughout the project in updating and exploiting the IPR strategy as results arise.

In the remainder of this document, each part of the IPR strategy is described in order to give a clear point of view of the plan to be followed regarding IPR.

[1] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

SUBJECT MATTER	TRADEMARKS	PATENTS	INDUSTRIAL DESIGN	UTILITY MODEL	TRADE SECRETS	COPYRIGHT
Invention		N/A		N/A		
Software (SW)		N/A			X	X
Scientific Article						X
Design of a Product	X		X			X
Name of a Product/Service	X					
Know-How					X	
Website	X		X			X

Table 1 Illustration of results protection

2.1 IP OWNERSHIP

As established in the GA and the CA, the mutual obligations arising from the joint activity of at least two project partners has been described and all the required actions to be done listed.

Any kind of Joint Agreement between partners will, if necessary, be arranged in order to enforce project results. The correct application of IPR policies will be guaranteed by the Innovation Manager in this regard.

2.2 ACCESS RIGHTS TO BACKGROUND AND RESULTS

Access Rights means the rights to use Results or Background under the terms and conditions laid down in the LEXIS GA and CA. Specifically, the purpose of the LEXIS CA is to specify, with respect to LEXIS project, the relationship among the partners, in particular concerning the organisation of their work, the management of project's activities and their rights and obligations relating to, *inter alia*: liability, access rights and dispute resolution.

For the sake of clarity, the definition of "Background" and "Results" are recalled as follows:

- "Background" means any and all data, information, know-how (tangible or intangible) whatever its form or nature, including any IPR s that are:
    - owned by a LEXIS partner or that a LEXIS partner has a right to license, prior to the effective date of LEXIS CA, or
    - developed or acquired by a LEXIS partner independently from the work in LEXIS even if in parallel with the performance of LEXIS.
- "Results" means any tangible or intangible output of LEXIS such as data, knowledge and information whatever their form or nature, whether or not they can be protected, which are generated in LEXIS as well as any rights attached to them, including Intellectual Property Rights.

For further details about Access Rights, reference should be made to Section 9 of LEXIS CA.

2.3 CONSORTIUM ORGANIZATION & FUND DISTRIBUTION

At the beginning of the project implementation, the consortium has signed the CA that defines the most important issues, e.g., fund distribution, reporting mechanisms with regular risk assessment and revised the management roles. It also defines the implementation of project specific issues including IPR, mostly occurring in WP1 and WP9.

During the quarterly reporting, the coordinator regularly collects information from all partners and updates it at the Participant Portal.

All the key roles of the project management (Innovation manager, Dissemination manager, Quality manager) were appointed; the communication tools were set; and the formal review procedure of the project outputs was determined.

Also, an External Advisory Board (EAB) consisting of representatives of external research/academia institutions not involved in the project consortium was established at the beginning of the project. The main goal of the EAB is to give informed feedback to the work of the project from an independent perspective and to help promote the LEXIS project within their domain(s).

2.4 IPR MANAGEMENT DURING THE PROJECT

IPR Management, as agreed in the CA, is led by the Innovation Manager and has the role of coordinating, managing and reporting all the activities and negotiations surrounding IPR issues such as patents, patent applications or other aspects of the IP domain.

To ensure a good level of awareness and results, the Innovation Manager will be in cooperation and coordination with the task leader of exploitation activities and actions.

During the project lifetime, IPR protection will be duly analysed and will be adequately followed and, if will come up, reported on the EU portal.

All the applications for IPR protection will contain, as a general behaviour, reference to EU funding set out as agreed in the GA and in the next version of this deliverable.

If some beneficiaries will own results jointly, these results will be detailed in the next version of the deliverable detailing the joint results and the indication of the participation of each partner. The behaviour is to follow the process indicated in the IPR guidelines for this kind of results.

To ensure a high qualitative level of the IPR documentation, the Quality Manager will check adequately all the documentation coming out and related to IPR, as final validation.

2.5 TRANSFER OF RESULTS

Results of LEXIS and their transfer are a crucial part of the IPR management strategy. Starting from the definition of the Ownership of Results (Full ownership, joint ownership) the LEXIS Consortium will define the strategy and rules of the Transfer of Results following the GA and the CA notwithstanding the recommendations of the EU as described in the MCARD 2020.

This is of critical importance keeping in mind that the life and impact on Research, Industry and Society of LEXIS beyond the end of the project will be determined by a clear and effective policy. On top of the question of the Transfer of Results between Parties involved and potentially their Affiliated Entities, the question of the Dissemination of Results and furthermore the Access Rights to Results, published or unpublished are linked.

A potential exploitation of LEXIS results will be defined first by the choices made in this regard by the LEXIS Consortium, with at stake the respect of Open Access to Scientific Publications, Research Data Rules on one hand and on the other hand a potential exploitation for commercial purposes where protections, restrictions, licensing rights may be required in some extend, but still need to be refined.

Results and their transfer have to be handled:

- During the project itself, including the 3 Pilots,
- Beyond the end of the LEXIS project.

The question about how to handle the Transfer of Results has to be answered taking in consideration:

- The framework as mainly defined by:
    - The EU Guidelines on Data Management in H2020,
    - The Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research data in H2020,
    - The GA,
    - The CA.
- The rules set-up for the access to the Background,
- IPR Management Rules during the Project and beyond,
- Open Sources and Standards,
- Specificities linked to each of the 3 Pilots,
- Definition of what the Consortium intends to achieve to manage the outcome of LEXIS and its management beyond the end of the LEXIS project,
- Regulations about exportations (restrictions linked to defence, geographies and other criteria) from the EU and potentially other governmental or international bodies (including NATO as an example).

As a consequence, the Transfer of Results has to be defined at a later stage in the LEXIS project. The definition of how to handle the "after-life" of LEXIS will be fundamental in this approach, as it should be anticipated before the end of LEXIS.

Legal advice from experienced specialists could be helpful in this complex subject, and the LEXIS consortium will seek for assistance from the European IPR Helpdesk where necessary.

Taking in consideration the diversity of our targets, for the Open Call and beyond the end of LEXIS, we have to anticipate various scenarios:

- Research Laboratories and Universities (public sector - Not for profit),
- Private sector - general rules,
- Private sector - start-up and SMEs specific cases.

2.6 OPEN SOURCE AND STANDARDS

Open Source technology is important for LEXIS: to achieve impact in the large HPC sector, it is important to have solutions which can be easily adopted by other centres. Standards are of interest to LEXIS to employ best practices in issues such as data management, data cataloguing, data indexing and data representation, but LEXIS is not specifically driving any standards activity throughout the course of the project.

LEXIS has a significant Open Source dimension both in terms of the applications running on the LEXIS platform as well as components which constitute the LEXIS platform. Example of Open Source technologies which existed before LEXIS and have a role within LEXIS are listed in Table 2, including their existing licenses and how they will be used within LEXIS. The final table will be included in the next version of the deliverable in M30.

TECHNOLOGY	EXISTING LICENSE	NOTE ON USE WITHIN LEXIS
iRods[2]	Permissive BSD license[3]	iRods will be used as the basis for data distribution, replication and staging within WP3
Ystia/Yorc[4]	Apache 2.0 License[5]	Ystia/Yorc will be used as the basis for deploying workload to the federated cluster driven by WP4 - LEXIS will contribute to the development of Ystia/Yorc
Cyclops[6]	Apache 2.0 License[7]	Cyclops will form the core of the LEXIS accounting and billing solution and will be developed further within LEXIS
TsunAWI[8]	GPL v2 with a small modification	TsunAWI is being used within WP6 to perform modelling of ocean movements and Tsunamis in particular
HEAppE Middleware9	GPL v3.0. for research purposes	HEAppE will be used as secure layer between Ystia/Yorc and HPC/Cloud infrastructure for the managing of computational task
FDB	Modified Apache 2.0 License	ECMWF's open-source FDB will be used as a back-end for the Weather and Climate Data API (WCDA)

Table 2 Open Source technology components used within LEXIS and their associated licenses

As well as pre-existing Open Source technologies, LEXIS will produce new software - in particular, the portal to be developed within WP8 will be Open Source such that it can be employed by other HPC centres; other new Open Source technologies may arise from some of the work within the pilot WPs.

The specific means for exploitation of Open Source technologies within LEXIS is still under discussion and the project is developing a process for articulating this; as would be expected, it is non-trivial, needing to incorporate the following:

- A comprehensive list of Open Source projects to which LEXIS will contribute,
- A list of LEXIS consortium members that will contribute to these projects,
- Commercial or otherwise agenda of these contributors,
- How pre-existing licenses can conflict with commercial agendas of partners,
- How to mitigate issues with combining licenses in software solutions comprising of a substantial number of constituent components,
- Routes to exploitation of Open Source work, particularly within the European HPC sector but also more globally.

This process is being led by the Innovation Manager and will be documented in detail in D9.6 - Report on IPR Management [1].

[2] https://github.com/irods/irods
[3] https://github.com/irods/irods/blob/master/LICENSE
[4] https://github.com/ystia/yorc
[5] https://github.com/ystia/yorc/blob/develop/LICENSE
[6] https://github.com/Cyclops-Labs/cyclops
[7] https://github.com/Cyclops-Labs/cyclops/blob/master/README.md
[8] https://swrepo1.awi.de/projects/tsunawi/
[9] http://heappe.eu

Standards activity within LEXIS will focus on two specific areas:

- Consistency with standard practices for securing access to HPC resources, in particular developing a solution which is compatible with ISO 27001.
- Using best practice standards or de facto standards for data management, data indexing and data representation. Further, for all data sets that are published within LEXIS, standardized data formats will be employed where possible.

3 INITIAL DATA MANAGEMENT PLAN

The Data Management Plan (DMP) specifies what data will be generated in the project and what data will be exploited and/or shared/made accessible for verification and reuse and how this data will be maintained. The DMP will evolve during the project to present the project progress in terms of data management. The LEXIS project opted in to the Open Access Research Data Pilot. The policy reflects the LEXIS CA regarding data management; it is consistent with the exploitation and protection of results.

An initial evaluation of the existing hardware and data-management-system infrastructure at the HPC sites was performed in Task 2.1 (Infrastructure Evaluation and Key Technology Identification for LEXIS) of WP2. The evaluation focused on existing data management systems and data-transfer interfaces, existing AAI, existing HPC/HPDA systems with the potential for use in LEXIS, and resilience of existing hardware with respect to data loss due to hardware failure or power outage. The detailed description of the analysis results is provided in Deliverable D2.1, Section 4 [2].

Also, the analysis of the pilot requirements focused on software applications and workflows has been done within Task 2.1 (WP2). The analysis results presented in Deliverable D2.1, Section 2 [2] also include requirements to data system infrastructure like storage, required bandwidth, etc.

Section 3.1 is focused on data management policies of the supercomputing centres whose infrastructure plays an important role during the implementation phase of the LEXIS platform.

The LEXIS project aims to build an advanced engineering platform at the confluence of HPC, Cloud and Big Data which will leverage large-scale geographically-distributed resources from existing HPC infrastructure, employ Big Data analytics solutions and augment them with Cloud services. Due to this reason, Section 3.2 describes the LEXIS platform data management overview including a workflow orchestration system, data management through the devised LEXIS data system, and a front-end portal designed for getting access to the computing resources and data leveraging LEXIS platform components.

The last three sections contain data management plans for the LEXIS pilots: Aeronautics pilot (Section 3.3), Earthquake & Tsunami pilot (Section 3.4), and Weather and Climate pilot (Section 3.5). Each section covers the data summary, Findable, Accessible, Interoperable and Re-usable (FAIR) data policy, data security, and plan for the data outputs.

Significant project results in the form of data collections and benchmarks will be archived in long-term repositories like Zenodo[10] or GitHUB[11] (if the data collection is closely connected with software). Additionally, institutional repositories (for example of IT4I or LRZ) will be used to make available appropriate documentation or relevant publications.

Within the course of the project, it is envisaged that the LEXIS Distributed Data Infrastructure (DDI), based on iRODS and EUDAT[12] B2SAFE, may serve as a full-fledged research data repository. Not being meant as a competitor to platforms like Zenodo or GitHUB, it serves the special purpose of an efficient distributed data management within

[10] ZENODO - https://zenodo.org
[11] GitHUB - https://github.com
[12] EUDAT - https://eudat.eu

LEXIS. It will, however, also incorporate components to ease the task of "making data FAIR" for LEXIS platform users directly (even when not uploaded e.g. to Zenodo). An envisaged connection to B2HANDLE ensures that data are given a PID; EUDAT B2FIND ensures that data can be successfully searched for. Data access is provided via the LEXIS web portal, and Interoperability and Reproducibility are fostered by a common metadata standard (apart from measures within the pilot use cases themselves), as we will describe in the following.

Metadata for public data sets

Data collections provided by the LEXIS project will meet the requirements for the FAIR data policy. To make them discoverable, the metadata will be set for each data collection. The LEXIS outputs will follow international standards, for example DataCite[13]. The approach is minimalistic metadata, helping to find the data, see what it is about, and know whom to talk to. Such a standard would be appropriate also as a minimal metadata standard for LEXIS. DataCite contains - apart from mandatory fields - several recommended/optional fields which could be specified as "mandatory in the LEXIS project" if needed; also, metadata sets following the DataCite standard can refer to additional domain-science/engineering metadata via the related Identifier field (specifying e.g. a URL for the additional metadata).

DataCite has 6 mandatory principal fields, which are:

- `Identifier` (normally a DOI, can be left empty and will be filled when DOI is requested),
- `Creator` (main researchers involved in producing the data),
- `Title` (of the data set),
- `Publisher` (entity which holds/archives/publishes/produces data),
- `PublicationYear` (year when data became/become/will become public; if this is not available, use the date preferred from a citation perspective),
- `ResourceType` (e.g. data set of weather-simulation data).

Each of these fields contains very few subfields for further description. Some fields can be specified multiple times (e.g. `Creator` if there were more than one main researcher).

In the Munich research data management context (University Libraries and LRZ), a recommendation on "how to use DataCite" has been developed, which has the purpose exactly of making data from different disciplines (humanities, engineering, etc.) as FAIR as possible as smoothly as possible. It requires the Subject field (scientific discipline, keywords) in addition to the fields mentioned above.

Further noteworthy recommended/optional fields[14] contain, e.g., further contributors, access/usage rights and related identifiers. LEXIS partners can use also for example extensions like `RelatedIdentifier` to link to documentation on a data set; or the `Rights` property. DataCite can also be used in JSON.

3.1 SUPERCOMPUTING CENTRES DATA MANAGEMENT POLICIES

3.1.1 IT4I

Human roles and administration process

IT4Innovations System Administrators are full-time internal employees of IT4Innovations, department of Supercomputing Services. The system administrators are responsible for safe and efficient operation of the computer hardware installed at IT4Innovations. Administrators have signed a confidentiality agreement.

User access to IT4Innovations supercomputing services is based on projects — membership in a project provides access to the granted computing resources (accounted in core-hours consumed). The project will have one **Primary**

[13] DataCite - https://en.wikipedia.org/wiki/DataCite, https://schema.datacite.org/meta/kernel-4.2/index.html
[14] Metadata - recommended/optional fields: https://schema.datacite.org/meta/kernel-4.2/doc/DataCite-MetadataKernel_v4.2.pdf, Table 4

Investigator, a physical person, who will be responsible for the project, and is responsible for approving other users access to the project. At the beginning of the project, Primary Investigator will appoint one Company Representative for each company involved in the project.

Company Representatives will be responsible for approving access to **Private Storage Areas** belonging to their company. Private Storage Areas are designated for storing sensitive private data. Granting access permissions to a Private Storage area must be always authorized by the respective Company Representative AND Primary Investigator. Available on request.

Users are physical persons participating in the project. Membership of users to LEXIS project is authorized by Primary Investigator. Users can log in to IT4Innovations compute cluster, consume computing time and access shared project storage areas. Their access to Private Storage Areas is limited by permissions granted by Company Representatives.

User data in general can be accessed by:

1. IT4Innovations System Administrators.
2. The user, who created them (i.e. the UNIX owner).
3. Other users, to whom the user has granted permission and at the same time have access to the particular Private Storage Area (in the case of data stored in the Private Storage Area) granted via the "Process of granting of access permissions" process.

Process of granting of access permissions

All communication with participating parties is in the manner of signed email messages, digitally signed by a cryptographic certificate issued by a trusted Certification Authority. All requests for administrative tasks must be sent to IT4Innovations HelpDesk. All communication with HelpDesk is archived and can be later reviewed.

Access permissions for files and folder within the standard storage areas (HOME, SCRATCH) can be changed directly by the owner of the file/folder by respective Linux system commands. The user can request HelpDesk for assistance on how to set the permissions.

Access to Private Storage Areas is governed by the following process:

1. A request for access to Private Storage Area for given user is sent to IT4Innovations HelpDesk via a signed email message by a user participating in the project.
2. HelpDesk verifies the identity of the user by validating the cryptographic signature of the message.
3. HelpDesk sends a digitally signed message with request of approval to the respective Company Representative and to the Primary Investigator.
4. Both the Company Representative and the Primary Investigator must reply with a digitally signed message with explicit approval of the access to the requested Private Storage Area.
5. System administrator at HelpDesk grants the requested access permission to the user.

Company representative or Primary Investigator can also send a request to HelpDesk to revoke access permission for a user.

Data storage areas

There are four types of relevant storage areas: HOME, SCRATCH, BACKUP and PRIVATE. HOME, SCRATCH and BACKUP are standard storage areas provided to all users of IT4Innovations supercomputing resources (file permissions apply). HOME storage is designed for long-term storage of data and is archived on the tape library - BACKUP. SCRATCH is a fast storage for short- or mid-term data, with no backups. PRIVATE storages are dedicated storages for sensitive data, stored outside the standard storage areas.

HOME storage

HOME is implemented as a two-tier storage. The first tier is disk array and the second tier is a NL-SAS disk array together with a partition of T950B tape library. Migration between the two tiers is provided by SGI DMF software.

DMF creates two copies of data migrated to the second tier: one to NL-SAS drives and the second on LTO6 tapes for backup.

HOME is realized on CXFS file system by SGI. Access to this file system on the cluster is provided by three CXFS Edge servers and a pair of DMF/CXFS Metadata servers, which export the file system via NFS protocol.

Each user has a designated home directory on the HOME file system at /home/username, where username is login name given to the user. By default, the permissions of the home directory are set to 750, and thus it is not accessible by other users.

SCRATCH storage

SCRATCH is running on a parallel Lustre filesystem with fast access. The SCRATCH filesystem is divided into two areas: WORK and TEMP.

1. WORK filesystem. Users may create subdirectories and files in directories /scratch/work/user/username and /scratch/work/project/projectid. The /scratch/work/user/username is private to user, much like the home directory. The /scratch/work/project/projectid is accessible to all users involved in project projected.
2. TEMP area. In this area, files that are not accessed for more than 90 days will be automatically deleted. Users may freely create directories in this area, and are fully responsible for setting correct access permissions of the directories.

PRIVATE storage

In order to provide additional level of security of sensitive data, we will setup dedicated storage areas for each company participating in the project. PRIVATE storage areas will be setup in a separate storage and will be not accessible to regular IT4Innovation users. IT4Innovations can additionally provide encryption of PRIVATE storage; the particular solution will be discussed with regards to security and performance considerations.

BACKUP storage

Contents of HOME storage are automatically backed up to tape library. There is a minimal period of retention, but no maximal, so we cannot guarantee time when the backups are removed from the tapes.

PRIVATE BACKUP storage

It is possible to setup dedicated backups of PRIVATE storage. In this case, unlike with the regular BACKUP, we can guarantee secure removal of data archived in PRIVATE BACKUP.

Data access

Physical security

All data storage is placed in a single room, which is physically separated from the rest of the building, has a single entry door and no windows. Entry to the room is secured by electromechanical locks controlled by access cards with PINs and non-stop alarm system. The room is connected to CCTV system monitored at reception with 20 cameras, recording and backup. Reception of the building has 24/7 human presence and external security guard during night. Reception has a panic button to call a security agency.

Remote access and electronic security

All external access to IT4I resources is provided only through encrypted data channels (SSH, SFTP, SCP and Cisco VPN).

Control of permissions on the operating system level is done via standard Linux facilities – classical UNIX permissions (read, write, execute granted for user, group or others) and Extended ACL mechanism (for a more fine-grained control of permissions to specific users and groups). PRIVATE storage will have another level of security that will not allow mounting the storage to non-authorized persons.

Data lifecycle

1. **Transfer of data to IT4Innovations:** User transfers data from his facility to IT4Innovations only via safely encrypted and authenticated channels (SFTP, SCP). Unencrypted transfer is not possible.
2. **Data within IT4Innovations:** Once the data are at IT4Innovations data storage, access permissions apply.
3. **Transfer of data from IT4Innovations:** User transfers data from to facility from IT4Innovations only via safely encrypted and authenticated channels (SFTP, SCP). Users are strongly advised not to initiate unencrypted data transfer channels (such as HTTP or FTP) to remote machines.
4. **Removal of data:** On SCRATCH file system, the files are immediately removed upon user request. However, the HOME system has a tape backup, and the copies are kept for indefinite time. We advise not to use HOME storage if you do not wish to keep copies of your data on tapes. PRIVATE storage will be securely deleted upon request or when the project ends.

Data in a computational job life-cycle

When a user wants to perform a computational job on the supercomputer the following procedure is applied:

1. User submits a request for computational resources to the job scheduler
2. When the resources become available, the nodes are allocated exclusively for the requesting user and no other user can login during the duration of the computational job. The job is running with same permissions to data as the user who submitted it.
3. After the job finishes, all user processes are terminated and all user data is removed from local disks (including ram-disks).
4. After the clean-up is done, the nodes can be allocated to another user, no data from the previous user are retained on the nodes.

All Salomon computational nodes are disk-less and cannot retain any data.

There is a special SMP server UV1 accessible via separate job queue, which has different behaviour from regular computational nodes: it has a local hard drive installed and multiple users may access it simultaneously.

ISO certification

IT4I has established and continually improves an internationally recognized information security management system, manages risks, and has established processes and regulations to secure information against misuse, unauthorized changes, and loss. In December 2018, IT4I was awarded ISO 27001 certification (ISO/IEC 27001:2013, ČSN ISO/IEC 27001:2014). The certificate was obtained for provision of national supercomputing infrastructure services, solution of computationally intensive problems, performance of advanced data analysis and simulations, and processing of large data sets.

3.1.2 LeiBniz Rechenzentrum (LRZ)

Human roles and administration process

In LRZ, system access is implemented in the scope of a three-layer hierarchical system. **Administrative rights** in the LRZ Identity Management / Access / Accounting system (”LRZ-SIM”) are held by heads of groups/departments and the user administration, and partially by LRZ’s employees, which are undergoing legal screening before employment. Below these, there is the community of **Master Users,** who can administer a **LRZ** (computing and/or data) **project**, i.e. add user accounts within the project and give the project users usage rights for LRZ systems the project is entitled to via the LRZ identity-management portal[15]. Projects have to be requested by heads of university organisational units (chairs, institutes) as Primary Investigators - who then decide who will be master user - or depend on successful proposals in the case of PRACE-Tier-0 HPC access. The lowest level is made up by **ordinary LRZ users**, who are only entitled to change a few personal settings (e.g. the password or their certificate DN - but *not* their account name and/or personal name and affiliation).

[15] https://idportal.lrz.de

User data on LEXIS members can - in this scheme - be accessed by the users themselves, by LEXIS master users (LEXIS LRZ staff and LRZ administrators (persons with administrative rights).

Users receive, when setting their password, the first time - i.e. before usage of LRZ systems - information on AUPs and export-control regulations; master users receive an extended introduction to the regulations and to their role.

System isolation vs. Sharing

LRZ systems are **generally intended for academic and not-strictly-confidential usage scenarios,** and users should generally be aware to configure their storage spaces etc. appropriately so that it is not readable for other users - when being on a shared storage system of a HPC cluster. For the LRZ Compute-Cloud and VMWare-Compute-Instance Services, LRZ makes a best effort to isolate users from one another.

The Data Science Storage (DSS) - see below - implements an extended user-administration concept to enable data sharing and with high flexibility, with permissions directly administrated by the responsible person for the data.

Data storage areas and rights management

HPC storage: $HOME, $WORK, $SCRATCH

LRZ uses, similar to IT4I, a 3-tier HPC storage concept, where space assigned to a project (by means of quota) on a HPC system increases in the line $HOME -> $WORK -> $SCRATCH and data safety decreases in the same line ($HOME backed up, $WORK not backed up, $SCRATCH: high watermark file deletion possible). These storage systems are separate for each cluster and provide up to several hundred TBytes of storage per user (on $SCRATCH). Rights management on these file systems is handled by standard unix permission. Each user is required to check and, if needed, manually adapt the umask if he wants to prevent readability of files by other users in the project.

DSS

The DSS of LRZ - the primary system for storage of massive amounts of data which have to be available on LRZ's HPC systems as well as the cloud - uses a more flexible access-control concept. Mounting the DSS via NFS exports is possible within practically all LRZ systems, but the IP addresses of the target systems have to be explicitly registered. Within a DSS space - having the form of a Unix project - it is possible to make subdirectories, also called "containers". LRZ-SIM users can then be "invited" (i.e. allowed access) to a container via an access-control-list mechanism. Within a container, users regulate access of other users (as far as allowed to the container) via Unix access rights. Nonetheless, on user request, containers provide a clear and efficient way of isolation.

TAPE

LRZ provides long-term tape storage on an IBM Spectrum Scale system, where users authenticate separately via node names and node passwords. In archive mode, data is mirrored in the Computing Centre of the Max Planck Society (MPCDF).

Metadata Handling and FAIR Research Data Management

Recently, LRZ has begun to develop a metadata-handling system called Let the Data Sing (LTDS) and thus aims to support basic FAIR-data services based on DataCite metadata and indexing for data sets e.g. on the DSS. If the approach is successful, in 2020 DOI-based publication of data sets will be supported as well as metadata export to the German research data portal GeRDI[16] and prospectively EUDAT.

Data access and sharing

Data sharing and downloading at LRZ traditionally work – besides scp/sftp-based mechanisms – with GLOBUS or GCT-GridFTP. Via GLOBUS sharing, data can be shared with other GLOBUS users, while GridFTP merely serves as a certificate-based, efficient mechanism to retrieve data.

[16] www.gerdi-project.org

Data lifecycle and responsibility for data

Data Lifecycle Support at LRZ comprises all steps such as data transfer (GridFTP/GLOBUS), massive data generation and processing (HPC systems), data pre-/postprocessing and smaller computational tasks (Compute Cloud) and data archival (Spectrum Protect Tape Archive). Data deletion is possible on file systems; backups and archive data can, however, normally not be deleted by users, which may be a concern in the case of (also "accidentally backed up") personal data. Upon ending of LRZ projects, users have the possibility (in case of Tier-0 HPC systems the explicit right without additional costs) to negotiate with LRZ a long-term on-site archival solution (data-only project on disk/tape for Tier-0 HPC systems, or special tape-backup node).

The responsibility for keeping all important data, data safety and security on the user level - as far as not guaranteed by the systems themselves - lies with the individual users generating the data on LRZ systems. If an user leaves a project, the data are handed over - with all rights and responsibilities - to the respective master user(s) of the project; if these are unavailable, LRZ administrators take (limited) responsibility.

DMPs at LRZ

In general, LRZ does not demand any DMP from its customers. LRZ is an infrastructure provider, and does not see its role in imposing any methodological or management rules on its customers, which often already have strong obligations to lay out their project planning, DMPs etc. to their funding agencies.

However, the LRZ DSS, which is the primary LRZ storage for long-term storage of scientific results (simulation, processed measurement data, etc.) for the years to come, is an exception here. With the acquisition of DSS quota, implying that the customer has to pay per GB of quota acquired, LRZ demands a very simple DMP. This DMP makes sure that the customer e.g.:

- Does not plan to flood the file system with too many files per folder, making folder access ineffective,
- Does not demand unrealistic bandwidths, especially when acquiring little amounts of storage (the more storage is acquired, the more systems LRZ can buy for writing in parallel and thus increasing the bandwidth),
- Does not plan to acquire DSS space while other systems would be more appropriate (e.g. file systems for temporary supercomputing storage),
- And has at least a rough idea about DMPs, data handling and data sharing, as well as FAIR concepts (but does not oblige him to implement them 1:1).

For customers having an obligation to write a DMP for their funding agency, the LRZ-DSS DMP has to be filled in as well, but is likely to be a subset of the DMP already created for the funding agency (except for e.g. a specific bandwidth question). The idea is to not create hurdles for using LRZ-DSS; if a customer has core-hours granted on LRZ's PRACE-Tier-0 machines, he can allocate up to 200TB per million of core hours granted by filing in a very much reduced query form instead of a DMP. LRZ-DSS Data Management Plan Template is presented in Appendix A.

ISO certification and data-safety regulations

LRZ is currently undergoing ISO 20000 and 27001 certifications, ensuring quality control of its systems including storage systems (however, it has to be emphasized that only the minority of LRZ's storage systems have long-term guarantees / SLAs, which is quite normal in scientific supercomputing centres).

LRZ I/SMS (Information Security Management System (ISMS) + Service Management System (SMS)) has been successfully audited for ISO 20000 and 27001 certifications in level 1 on date 10.05.2019. Level 2 audit will take place on 1 - 5 July 2019.

As part of the ISO process, mandatory security and data-safety measures partly already implemented earlier at LRZ (isolation of test from production systems, usage of SSL-encrypted protocols, email encryption in confidential or sensitive traffic) have been consistently formalised within a SMS and an ISMS.

3.1.3 ECMWF

ECMWF has rigorous data management plans related to its operational workflows, archival systems and delivery mechanisms however, these are not relevant to LEXIS. ECMWF's role as a data centre within LEXIS is different to LRZ's or IT4I's, as there is no provision for third-party access to HPC or data facilities — because ECMWF is a time-critical operational site.

The exception to this is the usage of the Weather and Climate Data API (WCDA), which will be hosted across all three data centres. The data management plan for the WCDA is discussed in Section 3.5 of this document, and in more detail in deliverable D7.1 [3].

3.2 LEXIS PLATFORM DATA MANAGEMENT OVERVIEW

The LEXIS platform will be built and implemented around different components taking care of authorizing users to get access to computing and storage resources, authorizing users to query available data, orchestrate complex work-flows, manage execution and monitoring of federated resources across distributed data centres, and providing billing functionalities according to used resources. To this end, LEXIS will leverage the following specific architectural components:

- A front-end portal (designed in WP8) integrating an Authorization and Authentication Infrastructure (AAI),
- An orchestration module integrating specific solution designed to interact with HPC clusters (WP4),
- A data system which will provide mechanisms for managing data along their lifespan (WP3).

Integration of all these components into a coherent federated system will be achieved through activities set up in a dedicated work package (WP2). This work package will ensure that all the platform sub-systems not only work together correctly, but also ensure data management policies will be applied accordingly. As such, specific and concrete data management policies will be defined during the project and reported in deliverables.

Orchestration (WP4)

An entire work package WP4 is devoted to the design and integration of a workflow orchestration system. The orchestration component will ensure the capability of scheduling workflow tasks also across distributed data centres (e.g. IT4I, LRZ) following a coherent resource federation mechanism. The tasks distribution will also be performed considering where data are generated (data locality), in order to avoid creation of bottlenecks that would reduce the overall performance and increase costs. To this end, data-related metrics (e.g. data volume, data location, etc.) helping define scheduling policies will be analysed during the project. Similarly, constrains brought by authentication and authorization process will be taken into account to correctly grant access to the data.

Data management (WP3)

Data management will be primarily managed through the devised LEXIS data system. WP3 is devoted to analyse and setting up a storage and management solution to this purpose. Figure 1 shows how this storage and management solution is organized, along with its integration with the orchestration module. Upon storage solutions for both Cloud and HPC- cluster portions of the LEXIS infrastructure (e.g. OpenStack Cinder will be used for managing Cloud portion — block storage service, parallel file systems will be part of the HPC clusters such as Lustre or GPFS), an entire set of APIs for querying, retrieving, filtering, getting data will be made available in the LEXIS platform. Also, integration with a monitoring solution will ensure (partial) data access tracking.

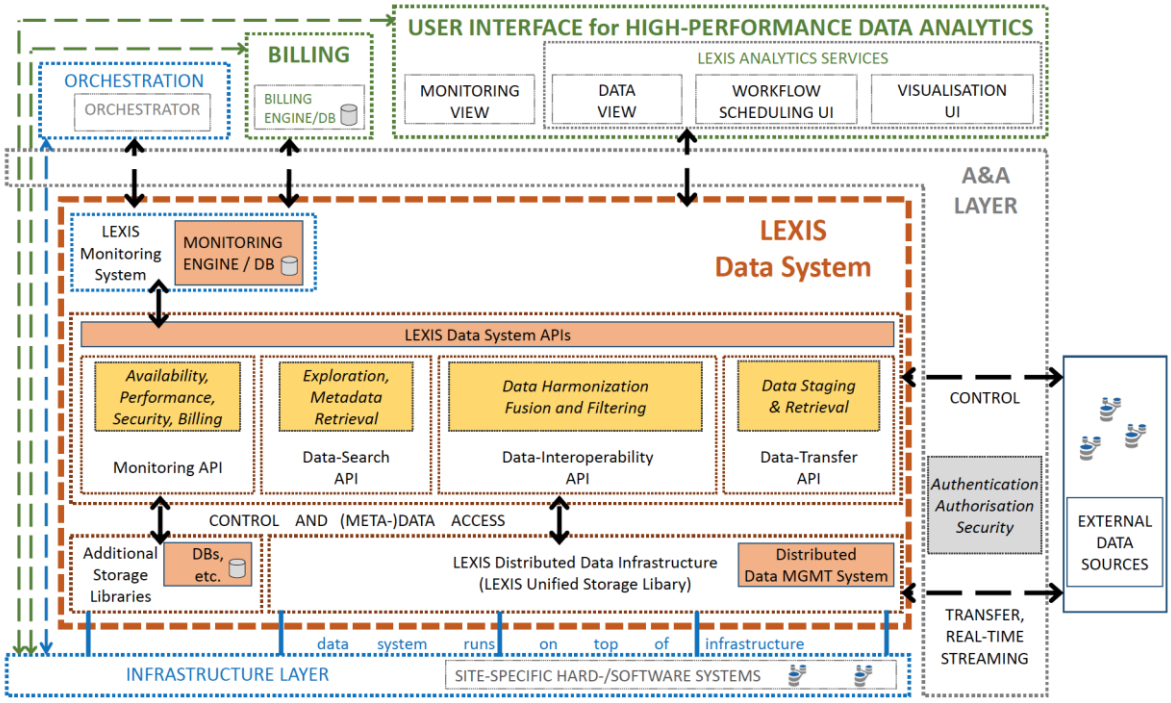

Figure 1 Data storage and management system

Such monitoring functionality will be implemented through activities carried out in WP8. LEXIS platform aims at exploiting advanced technologies to boost performance of executed workflows. To this end, data storage and management system will take advantage of the presence of Burst Buffer (BB) technology on some nodes in the LEXIS cloud layer. As such, BBs are special nodes, equipped with high-speed storage devices (i.e. NVMe, SSDs, etc.) which can be accessed transparently by workflow's tasks to speedup I/O intensive operations. To achieve this, the orchestration solution will be able to select and leverage such nodes to improve workflow performance. Similarly, data management policies will be defined, taking into account performance benefits and costs associated to the use of such resources. Also, acceleration for more basic functionalities, such as data compression, data encryption and data format adaptation will be part of the BB scope.

Portal (WP8)

Finally, another important component of the LEXIS platform is represented by the (front-end) portal. The development and integration of the portal with the remainder of the LEXIS platform is carried out in WP8. The portal aims to be the primary interface for third party to get access the computing resources (distributed across multiple computing centres and federated together) and data, as well as to enable the users to load their applications (workflows). As such, the portal will leverage all the other platform components, which expose their functionalities through specific APIs.

WP8 will gather data pertaining to users and organizations as they sign up to the portal. Consortium partners have considerable experience managing such data in an operational manner and standard GDPR compliant solutions will be employed which will be detailed in future deliverables.

3.3 AERONAUTICS PILOT DATA MANAGEMENT PLAN

In this section, the aeronautics pilot data management plan will be described.

3.3.1 Data Summary

The challenge of Avio Aero is to develop new technologies enabling the design and production of aeronautical components and modules for next generation greener and quieter engines. Identified tasks will focus on:

Multistage Low Pressure turbine modules simulation. By exploiting LEXIS technology, we intend to obtain a marked step-change: less time consuming HW/SW coupling, opening the doors to the "real time" design approach. Furthermore, the big data produced as a result will require proper solutions to be put in place for quick data access, management and post-processing.

A digital (r)evolution being applied to Gearboxes design and development that are a key actor of GE Avio business. The aim is to investigate CFD capabilities applied to simulate mechanical parts rotating in presence of air and oil. Nowadays, this kind of simulation is at the leading edge of numerical technology. Large computing time is deemed. Hence advanced Hardware (HW) solutions are envisaged and will open new scenery to mechanical parts optimization, vital for improving engines' performance and reliability.

These test cases will produce several simulations to test and validate novel LEXIS High Performance Computing (HPC) infrastructure developed as a primary objective of the project. Data collected will belong to four scenarios illustrated in Section 3.3.4, discriminating open and confidential/not shareable data:

1. Videos for Aeronautics Pilot cases (not subject to confidentiality restrictions),
2. Avio Aero Turbomachinery Confidential Data,
3. Avio Aero Rotating Parts Confidential Data,
4. Complementary dataset for open source CFD solver (not subject to confidentiality restrictions)

The ones related to real test cases (points 2 & 3) will be owned by Avio Aero, will be labelled as confidential and will not be shareable while the other ones, related to videos and data provided for CFD comparable analyses, will be of public domain and then re-usable. The category labelled as public has the objective to promote understanding and comprehension about which type of knowledge can be extracted thanks to Computational Fluid Dynamics Analysis (CFD): a great insight to explore complex flow fields and minimize loss sources in order to get a more efficient fluid process. Video files will be ".avi", ".mp4", or ".mpeg/.mpg" type while the size will be ranging from ten to hundred of MB's.

3.3.2 FAIR data

This section applies only to open Aeronautics data (with no IP restrictions) that gather input/output data or openly shareable results, related to Computer Aided Engineering (CAE) simulations carried out on Aeronautical products.

In order to make this open data FAIR, two categories of metadata have been identified:

- Engineering metadata,
- Computational metadata.

From the Engineering standpoint, the following mandatory metadata should be used:

- `Identifier`,
- `Creator`,
- `Title`,
- `Publisher` (e.g. Avio Aero),
- `PublicationDate`,
- `ResourceType` (e.g. CFD-simulation dataset).

In addition, the following other ones should be also taken into account to allow an effective data traceability:

- Product category (e.g. Turbine, Gearbox),
- CAE discipline (e.g. mechanical, fluid dynamic, thermal or multi-discipline),
- Used CAE solver,
- Description (free text),
- Other information (free text).

Finally, from the Computational perspective, the following metadata should be used to guarantee not only robust identifiability of data but also their proper reproducibility stating the same computational parameters and input file:

- Server,
- Software,
- Software version,
- Job name,
- Wall time,
- No. of CPUs,
- Memory required,
- Input file,
- Output directory.

Making data openly accessible

As explained above in Section 3.3.1, there will be both proprietary and public data.

The former data will be confidential to Avio Aero, since geometries and boundary conditions will be linked to real engines and industrial components in development and cannot be shared and disclosed.

On the contrary, the latter open/public data will serve to explain the big potential and results obtainable from CFD simulations and will be made available to public and third parties for the duration of the project on a dedicated platform, created by the LEXIS team. They will consist in some videos, showing specific features relating to turbomachinery or gears, and in simulations obtainable thanks to a free, open source CFD software, named OPENFOAM[17]. It has a large user base across areas of engineering and science, from both commercial and academic organizations. OPENFOAM has an extensive range of features to solve any complex fluid flows involving chemical reactions, turbulence and heat transfer, to acoustics, solid mechanics and electro-magnetics. Tutorials and user guide will be made available for public usage to promote a first approach to this type of analysis.

The IT4I team will work to perform this kind of simulation with OPENFOAM. Data from open literature turbo-machinery analyses will be most likely used as a priority.

Making data interoperable and increase data re-us

This applies only to open Aeronautics data related to the scenario #4 (Section 3.3.1).

[17] https://www.openfoam.com/

CAE data produced as output from the computer-aided simulations that the aeronautics pilots rely on are deterministic, so the reproducibility of output data is always guaranteed under the same computational parameters and input file.

The produced CFD results should be generated according to standard file formats or structures allowing other researchers, institutions, organisations, countries, etc. to re-use them as input for refining or executing simulations related to other CAE disciplines (such as mechanical one).

Moreover, the adoption of an open source solver will definitely contribute not only to further increase data re-usability but also to disseminate CFD importance and impacts on aeronautical products optimization.

3.3.3 Data Security

A list of IT security and compliance questions about several topics (Change Management, Data Protection, Endpoint Security, Identity & Access Management, Incident & Problem Management, Incident Detection and Response, IT Governance, Logging and Monitoring, Network Design and Operations, Physical & Environmental, Registration & Asset Management, Secure Development, Security Testing) has been defined with the aim of supporting the analysis of mechanisms to be set up for securing the LEXIS federated infrastructure.

Generally speaking, one point of contact responsible for data security of LEXIS data centres shall be required and physical access to the LEXIS infrastructure (including also network devices such as routers, switches, etc.) must be protected to allow access only by approved people. The list of authorized people must be documented.

Specifically, from the storage standpoint, all data shall be safely stored in the LEXIS data system infrastructure, that should be certified and should comply with formal standards, such as ISO 27001 or ISO 47000. Specifically, confidential data must be stored with logical access controls, e.g., individual named account access and password protection. Additionally, if stored on portable devices, information must be encrypted. All backup removable media (tapes, USB drives, etc.) must be clearly and properly externally labelled. Delivery or handling process of such media must be documented, and a security review done of it.

From the data protection standpoint, a documented application recovery plan must be implemented and should include at minimum the following elements:

- Application Recovery scope,
- Required Recovery Time Objective (RTO),
- Required Recovery Point Objective (RPO),
- Process to invoke the Application Recovery plan, including parties authorized to invoke the process for application recovery,
- Steps performed to validate that the application recovery has been completed successfully,
- Frequency of testing and validation,
- Minimum requirement of annual test,
- Annual review and sign-off of the Application Recovery plan and test results.

Moreover, in the case of data transmission, transit encryption is needed and shall meet the following requirements:

- Transmissions over the web must use cryptographic protocols,
- Users must use secure versions of network protocols. E.g. HTTPS instead of HTTP, SFTP/FTPS instead of FTP,
- Other non-web traffic must be encrypted using Avio Aero’s approved cryptography algorithms, modes and key sizes.

3.3.4 Plan of the outputs

Categories of the research outputs specifically related to WP5 activities are listed as follows:

1. Videos for Aeronautics Pilot (not subject to confidentiality restrictions),
2. Avio Aero Turbo-machinery Confidential Data,
3. Avio Aero Rotating Parts Confidential Data,
4. Complementary dataset for open source CFD solver (not subject to confidentiality restrictions).

For each category a specific table provides additional information.

ID	ITEM	DESCRIPTION
D1	**Dataset name and reference**	AA DIS OD (Avio Aero Dissemination Open Data) This dataset refers to illustrative videos based on open data about Turbomachinery and/or Rotating parts. Identifier - DOI (to be defined)
	Dataset description	This dataset refers to the computer-aided engineering simulations, illustrated through videos, to show external people what can be obtained from CFD investigation in terms of insights about flow structure in Turbomachinery and/or Rotating parts.
	Standards and metadata	Video files will be ".avi", ".mp4", or ".mpeg/.mpg" type. Metadata not applicable.
	Data sharing	This data can be openly accessed on a dedicated LEXIS platform, that will provide capabilities for only executing the video with no download, or through dedicated URLs on the websites or platforms of Avio Aero's SW providers allowing visualization only.
	Is dataset accessible?	Yes
	Is dataset reusable?	Not applicable
	Archiving and preservation (including storage and backup)	Throughout the whole duration of LEXIS project, this dataset will be archived at LEXIS data system infrastructure or on proprietary repository (owned by Avio Aero and/or SW suppliers involved in the project). The purpose will be focused on disseminating CFD importance and impacts on aeronautical products performance optimization.

Table 3 Aeronautics pilot: D1 - AA DIS OD

ID	ITEM	DESCRIPTION
D2	**Dataset name and reference**	AA TM CD (Avio Aero Turbomachinery Confidential Data) This dataset refers to the numerical investigations supporting the Aeronautics Turbomachinery Use Case carried out by Avio Aero in LEXIS project. Identifier - to be defined
	Dataset description	This dataset refers to the computer-aided engineering simulations that the Aeronautics Turbomachinery Use Case relies on, and includes:

		• the CFD input data (turbine geometry and boundary conditions) and produced output data (aerodynamic solution) of the multistage Low Pressure turbine to be investigated, • the results from this aeronautics case study in terms of KPI.
	Standards and metadata	Standards not applicable. For only internal purposes, the metadata described in Section 3.3.2 will be associated to this specific dataset.
	Data sharing	Due to IP constraints, this dataset: • Can be accessed only by the LEXIS partners that will contribute to WP5 operations, and exclusively if strictly needed for the purpose of executing the CAE simulations, • Cannot be used, disclosed to others, or reproduced, without the express written consent of GE Avio S.r.l. Data sharing only applies here to the results in terms of KPI's because neither the simulation I/O data nor the CFD solver (TRAF code owned by University of Florence is subject to NDA) used in this study can be shared. The results in terms of KPI's will be provided in the deliverable D5.5.
	Is dataset accessible?	No. Only results in terms of KPI's will be accessible.
	Is dataset reusable?	No
	Archiving and preservation (including storage and backup)	Because of IP constraints, the dataset will be temporary available at LEXIS data system infrastructure throughout the whole duration of LEXIS project, only for the purpose of executing numerical simulations supporting the Aeronautics Turbomachinery Use Case. From the archiving and preservation perspectives, no long-term repository is required to support needs and goals of this case study. The long-term storage and backup of this dataset will be implemented on the internal Avio Aero Digital Technology systems.

Table 4 Aeronautics pilot: D2 - AA TM CD

ID	ITEM	DESCRIPTION
D3	**Dataset name and reference**	AA RP CD (Avio Aero Rotating Parts Confidential Data) This dataset refers to the numerical investigations supporting the Aeronautics Rotating Parts Use Case carried out by Avio Aero in LEXIS project. Identifier - to be defined
	Dataset description	This dataset refers to the computer-aided engineering simulations that the Aeronautics Rotating Parts Use Case relies on, and includes:

		• The CFD input data (CAD mesh and flow particles modelling, boundary conditions) and produced output data (fluid-dynamics air-gas mixture solution) of the gearbox to be investigated, • The results from this Aeronautics case study in terms of KPI's.
	Standards and metadata	Standards not applicable. For only internal purposes, the metadata described in Section 3.3.2 will be associated to this specific dataset.
	Data sharing	Due to IP constraints, this dataset: • Can be accessed only by LEXIS partners that contribute to WP5 operations, and exclusively if strictly needed for the purpose of executing the CAE simulations supporting the Aeronautics Rotating Parts Use Case, • Cannot be used, disclosed to others, or reproduced, without the express written consent of GE Avio S.r.l. The data sharing only applies here to the results in terms of KPI's because neither the simulation input/output data nor the commercial CFD solver (*ALTAIR Nanofluidx code*) used in this study can be shared. The results in terms of KPI's will be provided in the deliverable D5.5.
	Is dataset accessible?	No. Only results in terms of KPI's will be accessible.
	Is dataset reusable?	No
	Archiving and preservation (including storage and backup)	Because of IP constraints, this dataset will be temporary available at LEXIS data system infrastructure throughout the whole duration of LEXIS project, only for the purpose of executing the CAE simulations supporting the Aeronautics Rotating parts Use Case. From the archiving and preservation perspectives, no long-term repository is required in LEXIS to support needs and goals of this case study. The long-term storage and data backup will be implemented on the internal Avio Aero Digital Technology systems.

Table 5 Aeronautics pilot: D3 - AA RP CD

ID	ITEM	DESCRIPTION
D4	**Dataset name and reference**	AA TM OD (Avio Aero Turbomachinery Open Data) This dataset refers to input data supporting Aeronautics Turbomachinery open demonstration. Identifier - DOI (to be defined)
	Dataset description	This dataset refers to open input data feeding CAE simulations to be run by IT4I on LEXIS platform based on an open source CFD solver (for example OpenFOAM). It will include also the results coming out from numerical simulations.

	Standards and metadata	I/O file formats. The metadata described in Section 3.3.2 and associated to this specific dataset will be openly available.
	Data sharing	These data can be openly accessed on a dedicated LEXIS platform that will provide data downloading capabilities.
	Is dataset accessible?	Yes
	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	Throughout the whole duration of LEXIS project, this dataset will be archived and available at LEXIS data system infrastructure. The aim will be focused on disseminating CFD importance and impact potentially driving Aeronautical products performance optimization.

Table 6 Aeronautics pilot: D4 - AA TM OD

3.4 EARTHQUAKE AND TSUNAMI PILOT DATA MANAGEMENT PLAN

The earthquake and tsunami pilot will rely on specific datasets to measure its progress and allow external entities to interact and exploit the pilot results. The principle is that those data sets are open, but some of the results are considered sensitive in the sense that they may be mis-interpreted, and therefore, authorities may restrict their availability to the general public. We have striven however to define publicly available data sets, while considering the question of possible restricted datasets, to cover all cases.

3.4.1 Data Summary

Data sets in this pilot are of two types: input datasets, that represent the knowledge of a geographic area and a scenario, and on which simulations are run. And output datasets, that contain the results of the simulations and processes of the pilot.

These datasets represent the baseline input and output data of the pilot; measuring the time needed to process and simulate (and produce the output data sets) are the key elements for measuring the proper execution of the pilot.

Data sets are in XML, de-facto standard vector formats (e.g. shapefile), grid and georeferenced TIFF formats. Some datasets have a public origin, such as OpenStreetMap data, SRTM digital elevation models or satellite data acquired in the framework of SEM (can be either public/open or commercial/restricted). The expected size of the data is, in total, about 100GB. Those datasets may be useful for disaster response state agencies and entities, and to technology provider to tackle the performance problems associated with the pilot constraints (beyond or complementary to what is undertaken in the project).

3.4.2 FAIR data

Making data findable, including provisions for metadata

The data will be stored on data repositories with digital object identifiers. We will choose in priority public repositories as long as they allow us to comply with the constraints on the datasets access. Datasets will all have a metadata description, and, in the case of datasets with access restrictions, their metadata will be publicly available. A semantic versioning scheme will be used to track versions of the datasets.

Making data openly accessible

Input datasets will be freely available, except if we are unable to find a source allowing public access to their high resolution digital elevation model data. Output datasets will be available freely by default, except in the rare cases where regulations force us to regulate access so as to ensure responsible use of the data contained in those sets. Data formats will be standardized, open formats as used by the common open source tools of the domain, and format information will be included in the metadata associated with each dataset. For restricted access datasets, the partner responsible for generating that data will be the point of contact for requesting an access to the data. We will use the standard metadata scheme.

Making data interoperable

The datasets will use common representation and formats used in the domain, and, as such, will have an implicit link to the domain ontologies (geographic data, earthquake data, etc.). For less common data, an effort will be made to extend the common ontologies for describing the relevant data set.

Increase data re-use (through clarifying licenses):

To increase data re-use, publicly available datasets (or datasets extracted from a public source through a process of selection and refinement) will carry the same license as their source. The pilot partners will search to establish a common license for datasets for which public access is restricted, reusing if possible a pre-existing licensing scheme. Apart from that control as a protection over misinterpretation or misuse of the result datasets, the data will be reusable for as long as storage can be ensured for it.

3.4.3 Data Security

The data is not sensitive to security issues and may be recreated. Recreating those datasets will probably result in slightly different datasets, since some of the input datasets represent a freeze at a certain point in time of a global dataset under continuous update. Meaning and importance of the dataset will remain the same, however.

3.4.4 Plan of the outputs

ID	ITEM	DESCRIPTION
D5	**Dataset name and reference**	OSMGlobalBaselineAndOneWeekUpdate DOI - (to be generated)
	Dataset description	OpenStreetMap global data set at a predefined date (T_0) and a weekly update on that global data set (T_0 + one week). OpenStreetMap global data set for benchmarking at a specific point in time.
	Standards and metadata	XML, OpenStreetMap metadata
	Data sharing	Free sharing. Licence: OpenStreetMap license
	Is dataset accessible?	Yes
	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	Data set is available at gfz obm storage: http://data.obm.gfz-potsdam.de/Lexis/ and will be available at long-term repository Zenodo.

Table 7 Earthquake and tsunami pilot: D5 - OSMGlobalBaselineAndOneWeekUpdate

ID	ITEM	DESCRIPTION
D6	**Dataset name and reference**	OpenBuildingMapBaseline DOI - (to be generated)
	Dataset description	OpenBuildingMap dataset based on OSMGlobalBaselineAndOneWeekUpdate.
	Standards and metadata	Shapefiles
	Data sharing	Free access with a reference to LEXIS project. Licence: OpenStreetMap derived
	Is dataset accessible?	Yes
	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	Data set is available at gfz obm storage: http://data.obm.gfz-potsdam.de/Lexis/ and will be available at long-term repository Zenodo.

Table 8 Earthquake and tsunami pilot: D8 - OpenBuildingMapBaseline

ID	ITEM	DESCRIPTION
D7	**Dataset name and reference**	TsunAWIBaselineMesh DOI - (to be generated)
	Dataset description	Mesh of the targeted area for TsunAWI, for full scale simulations.
	Standards and metadata	ASCII files for vertex coordinates, connectivity, water depth / topography, bottom roughness (optional).
	Data sharing	Free access (unless restricted by DEM source).
	Is dataset accessible?	Yes
	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	Data set will be available at Zenodo repository.

Table 9 Earthquake and tsunami pilot: D7 - TsunAWIBaselineMesh

ID	ITEM	DESCRIPTION
D8	**Dataset name and reference**	TsunAWIBaselineWaveHeight DOI - (to be generated)
	Dataset description	Wave height data as computed by TsunAWI.
	Standards and metadata	netcdf on full mesh, interpolated raster data
	Data sharing	Free access
	Is dataset accessible?	Yes

	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	Data set will be available at Zenodo repository.

Table 10 Earthquake and tsunami pilot: D8 - TsunAWIBaselineWaveHeight

ID	ITEM	DESCRIPTION
D9	**Dataset name and reference**	SEM output data - vector and raster DOI (When updated)
	Dataset description	Vector: Reference and crisis layer as vector file Raster: ready-to-print maps
	Standards and metadata	Vector: ESRI Shapefile, Google Earth KML/KMZ files Raster: PDF, JPEG
	Data sharing	Free access (unless Copernicus Emergency Management Service activations that have been declared as sensitive)
	Is dataset accessible?	Yes
	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	Copernicus Emergency Management Service activation data will be available here: https://emergency.copernicus.eu/mapping

Table 11 Earthquake and tsunami pilot: D9 - SEM output data - vector and raster

3.5 WEATHER AND CLIMATE PILOT DATA MANAGEMENT PLAN

3.5.1 Data Summary

In the weather and climate pilot, curated data management is crucial in order to allow interoperability between observational data sources, global- & regional-scale weather models, and specific impact models. So much so, that a specialist weather & climate data API (WCDA) is being created to ensure data is fully-managed from both an organisational and performance perspective. Most of the data related to the weather and climate pilot, including observational data and computational-model outputs, will be managed by the WCDA — with some non-curated data using the LEXIS generic Distributed Data Infrastructure (DDI) platform instead.

The data used, generated and exchanged by the weather and climate pilot is typically in industry-standard scientific formats such as GRIB, BUFR, ODB or NetCDF, and range from a few megabytes to several terabytes. For a much more detailed analysis of the types of data involved in the weather and climate pilot, and a preliminary design of the WCDA, see deliverable D7.1 [3].

3.5.2 FAIR data

The Weather & Climate Data API (WCDA) is designed to excel at making data findable, accessible, interoperable and re-usable (FAIR). It will achieve this by ensuring data is highly-curated, leveraging scientifically-meaningful metadata to provide robust and transparent indexing.

Making data findable, including provisions for metadata

The metadata schemas, used to access different data via the WCDA, will be available as part of the REST interface. Users will be able to query the WCDA API to probe the data sets available, from which requests for data can be constructed. Due to the use of scientifically-meaningful metadata, most data will be self-describing, making it easy to find relevant data once the metadata schema is known.

ECMWF has a long history of effective data curation, with a meteorological archive (MARS) containing over 200 petabytes of highly-curated weather and climate data. This experience will be used to guide the metadata creation and management process for data in the weather and climate pilot.

Making data openly accessible

As far as possible, data will be made public and exposed via the LEXIS portal, however, this will be limited in cases where data is proprietary — which especially concerns real-time results from numerical models or third-party data sources such as in-situ weather observations. In many cases, the data being used in the weather and climate pilot is already open and can be accessed directly (e.g. ECMWF's public datasets or Copernicus data).

Relevant mechanisms will be used for the purpose of sharing code and scientific reports arising from the weather and climate pilot. For example, we strive to make most of the software open source, and the LEXIS GitHub page will host this. Many of the components which will be used to build the WCDA are already open source, including ECMWF's FDB technology.

Making data interoperable

There are two key aspects of the data management plan which ensure interoperability of data in the weather and climate pilot. Firstly, industry-standard file formats such as GRIB, NetCDF, BUFR and ODB2 are used as far as possible, ensuring that data is universally serviceable. Where necessary, these file formats will be enhanced by relevant standards and procedures to ensure consistency across, and beyond, the pilot. Secondly, the highly-curated, scientifically-meaningful metadata and indexing strategy will ensure strict data management, facilitating robust interoperable workflows. For more information on the curation process, refer to deliverable D7.1 [3], Section 3.3. For data which is stored in the LEXIS DDI, rather than the WCDA, the structure and standards of the data will be documented appropriately.

Increase data re-use (through clarifying licenses):

As mentioned, much of the data relevant to the weather and climate pilot is proprietary model output, used at an intermediate stage of the pilot workflows, or third-party observations used as inputs, which cannot be shared publicly. However, the weather and climate pilot will make the outcomes of its workflows publicly available where possible. This includes:

- WRF simulation model output,
- RISICO forest fire risk model results,
- Continuum hydrological model results,
- NUM urban model results.

For full details, see Section 3.5.4 below.

3.5.3 Data Security

The WCDA and LEXIS DDI will both integrate with the LEXIS AAI system and allow dataset and user-group policies to control access to public and non-public datasets in the weather and climate pilot. In terms of data recovery, the underlying data storage for the WCDA will be hosted in a resilient and redundant storage to allow reasonable recovery. The LEXIS DDI also makes suitable arrangements for data recovery.

3.5.4 Plan of the outputs

ID	ITEM	DESCRIPTION
D10	**Dataset name and reference**	WRF simulation model output
	Dataset description	2D and 3D meteorological fields, over different locations (Italy, France and Europe at large) at hourly temporal resolution.
	Standards and metadata	NetCDF format, standardized for the WCDA
	Data sharing	Free Access
	Is dataset accessible?	Yes
	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	A moving time-window of data will be available, tracking the real-time forecast workflows. This will be available for the duration of the LEXIS project. Sample data may be archived for future research.

Table 12 Weather and climate pilot: D10 - WRF simulation model output

ID	ITEM	DESCRIPTION
D11	**Dataset name and reference**	RISICO forest fire risk results
	Dataset description	Fire rate of spread, Fire Danger Index, Fire Danger early warning
	Standards and metadata	NetCDF format, standardized for the WCDA
	Data sharing	Free Access
	Is dataset accessible?	Yes
	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	A moving time-window of data will be available, tracking the real-time forecast workflows. This will be available for the duration of the LEXIS project. Sample data may be archived for future research.

Table 13 Weather and climate pilot: D11 - RISICO forest fire risk results

ID	ITEM	DESCRIPTION
D12	**Dataset name and reference**	Continuum hydrological model results
	Dataset description	discharge time series in different cross sections, daily soil moisture map, daily evapotraspiration map
	Standards and metadata	NetCDF format, standardized for the WCDA
	Data sharing	Free Access
	Is dataset accessible?	Yes

	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	A moving time-window of data will be available, tracking the real-time forecast workflows. This will be available for the duration of the LEXIS project. Sample data may be archived for future research.

Table 14 Weather and climate pilot: D12 - Continuum hydrological model results

ID	ITEM	DESCRIPTION
D13	**Dataset name and reference**	NUM urban model results
	Dataset description	Hourly timeseries of 2D Airquality concentrations of NO2 and PM10.
	Standards and metadata	NetCDF format, standardized for the WCDA
	Data sharing	Free Access
	Is dataset accessible?	Yes
	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	A moving time-window of data will be available, tracking the real-time forecast workflows. This will be available for the duration of the LEXIS project. Sample data may be archived for future research.

Table 15 Weather and climate pilot: D13 - NUM urban model results

4 CONCLUSION

In this deliverable, that is the first issue of the IPR and Data Management Approach, the approach for handling the IPR strategy and the Data Management plan was set up.

All these aspects will be maintained and updated during the course of the project and in the next version at the end of the project and will show the achieved results in term of IPR protection and DMP; each section will detail the above described strategies and approaches in order to represent the effectiveness of LEXIS platform and pilots at the end of the project showing up also the strategies to be followed beyond the project.

REFERENCES

[1] LEXIS, *Deliverable, D9.6: Report on IPR Management.*

[2] LEXIS, *Deliverable, D2.1 Pilots needs / Infrastructure Evaluation Report,* 2019.

[3] LEXIS Deliverable, *D7.1 Architectural Requirements and System Design for Interchange of Weather & Climate Model Output between HPC and Cloud Environments.*

[4] https://iprhelpdesk.eu/Fact-Sheet-IP-Management-H2020-Project-Implementation-and-Conclusion.

A LRZ-DSS DATA MANAGEMENT PLAN TEMPLATE

Basic Information			
ID	Question	Your answer	**Example**
A1	**SuperMUC-NG Project:**	<Insert Text>	pr12ab
B2	**Principal Investigator:** **LRZ User ID** **Name** **Affiliation** **Address** **Email** **Phone**	<Insert Text>	di12xy01 Dr. Erika Mustermann Institut für Studien Technische Universität München Musterstr. 1 Erika.Mustermann@tum.de +49 89 12345678
B3	**Masteruser/Projectmanager:** **LRZ User ID** **Name** **Affiliation** **Address** **Email** **Phone**	<Insert Text>	di12xy01 Dr. Erika Mustermann Institut für Studien Technische Universität München Musterstr. 1 Erika.Mustermann@tum.de +49 89 12345678 **LRZ considers this person as the "owner" of the data !**
B4	**Describe or link to the Research Data Management policy and practices , which you will follow (including practices About Intellectural property)**	<Insert Text>	https://www.ub.tum.de/open-access-policy https://secure.pangaea.de/curator/files/pangaea-data-policy.pdf http://www.en.uni-muenchen.de/funktionen/privacy/index.html
B5	**What would be the impact if your application does not get accepted:**	<Insert Text>	We would have to cancel our CPUh application and look for another SC-Center that supports us. We would reduce the resultion of our model and thus produce a smaller amount of data that can be handled by MWN Storage. The result would be a severe decrease of scientific quality.
Technical Information			
T1	**Amount of storage space requested:**	<Insert Text> TB	500 TB Keep the unit TB!
T2	**Number of files to be stored:**	<Insert Text> million files	500000 files
T3	**Time period for the storage:**	from: <Insert Text, yyyy-mm-dd> to: <Insert Text, yyyy-mm-dd>	from: 2019-06-01 to: 2023-06-01
T4	**Bandwidth requirements, expected data transfer in 24-hours.** **Do you have special requirements regarding the bandwidth for data access?**	<Insert Text>	We require a (24h-average) of 6 MB/s, i.e., we expect an average transfer of 0.5 TByte/day. *Special requirements would arise if you plan to exceed a (24h-avg.) bandwidth of 9MB/s per TB storage space – available bandwidth scales with storage space for technical reasons.)*
T5	**From which systems do you plan to access the data on DSS:**	<Insert Text>	e.g., LRZ Linux-Cluster LRZ Compute Cloud HLRS XYZ JSC JXYZ
T6	**Why is SuperMUC-NG DSS the preferable data management solution in comparison with other services at LRZ or your home institution:** (Examples of other LRZ services include but are not limited to: Tape Archive, SCRATCH, WORK, HOME))	<Insert Text>	We cannot use MWN Storage, because... We cannot use tape, because...
T7	**Do you plan to enable regular backups of your DSS data (how often):**	<Yes/No>	Yes, twice a week *If yes additional costs may arise*
Data Sharing			
			Beispiel
D1	**For how long do you need exclusive use of the data and why**	<Insert Text>	We would like to have exclusive usage for 3 years, since we want to publish several papers.
D2	**Are there plans for sharing the data in collaborations:** If yes, please sketch the collaboration including information about the affiliation of the collaboration partners.	<Insert Text>	Yes, we will collaborate with the UK Data archive of transfusional Spongeness. While we are interested in the performance data, their interest is to analyze the sponge models and compare them with the empirical findings.
D3	**Describe the efforts you plan to make your data FAIR (Findable, Accessible, Interoperable and Reusable):** Consider commenting on the following aspects: • equipment of the data and metadata with persistent identifiers (PIDs) • metadata-standard(s) • registration to search indexes • plans to persist metadata after data deletion • qualified references to other (meta)data • community-specific-standards for data description • documentation of data provenance	<Insert Text>	We will annotate our data with DataCite (which includes DOIs as PIDs). Registered DataCite data are findable via the search index of DataCite. We will host a landing page to our data at a LRZ webserver which can be operated even after the data are deleted. The standards of our community are best described by this paper 10.12034/asasf7wbqw8bf.set At the moment there is no provenance tracking planned that goes beyond the method sections in the papers we want to publish. LINK
Ethics and Legal Compliance			
I			Beispiel
E2	**Do you want to make your data available? Under which license?**	<Insert Text>	Yes CC-BY 4.0
E3	**Do you want to store Personal Identifiable Information (PII):**	<Yes/No>	NO **(If YES, you must contact LRZ before !)**
E4	**Do you want to store special categories of PII (Art. 9 of EU General Data Protection Regulation or DSGVO Art 9):**	<Yes/No>	NO **(If YES, you must contact LRZ before !)**
E4	I accept the Terms and conditions for DSS: • https://doku.lrz.de/display/PUBLIC/DSS+Terms+and+Conditions • https://www.lrz.de/wir/regelwerk/richtlinien_filesysteme_HPC	Yes/No	YES