SERVICE REQUEST

Service Facility in Support of the Strategic Development of International Cooperation in Research and Innovation N°30-CE-0838742/00-87 [PP-04341-2016]

SF(16) "Advancing the Design of the South-East European International Institute for Sustainable Technologies (SEEIIST)"

Deliverable 4.7 A – Report on the regional potential, capacities, opportunities and add-on benefit for the region for introducing a Digital Hub

As of 13.08.2020
Deliverable 4.7 A

European Commission
Directorate-General for Research and Innovation
Directorate H: International Cooperation
Unit 05: Horizon Europe Association
Desk Officer: Bernhard Fabianek
1049 Brussels
Belgium

Request for Service for a Specific Contract on “Advancing the Design of the South-East European International Institute for Sustainable Technologies (SEEIIST)”

Processing number: Pursuant to the Framework Service Contract 30-CE-0838742/00-87 [PP-04341-2016]
Submission date: 13.08.2020

Contact persons:
Jonas Abs
Service Delivery Manager
Deutsches Zentrum für Luft- u. Raumfahrt e.V., DLR.
Address
Heinrich-Konen-Str. 1
53227 Bonn, Germany
Tel. +49 228 3821 2063
Fax +49 228 3821 1408
E-Mail jonas.abs@dlr.de

Dr. Olaf Heilmayer
INT Service Facility Coordinator,
Framework Contract Leader
Deutsches Zentrum für Luft- u. Raumfahrt e.V., DLR.
Address
Heinrich-Konen-Str. 1
53227 Bonn, Germany
Tel. +49 228 3821 1443
Fax +49 228 3821 1408
E-Mail olaf.heilmayer@dlr.de

service.facility@dlr.de
Table of Content

1. Introduction ........................................................................................................................................... 4
2. Scope & Areas of Research and Innovation in the Digital Hub ................................................................. 6
   2.1. The Scope ........................................................................................................................................ 6
   2.2. Research and Innovation .................................................................................................................... 8
3. Challenges ................................................................................................................................................ 10
4. Opportunities .......................................................................................................................................... 13
5. Principal Competencies of the Digital Hub .............................................................................................. 15
6. Staff ......................................................................................................................................................... 18
   6.1. Core Staff ....................................................................................................................................... 18
   6.2. Supporting Staff ............................................................................................................................... 18
   6.3. External Collaborators .................................................................................................................... 19
7. Regional Considerations .......................................................................................................................... 20
   7.1. Present Capabilities ......................................................................................................................... 20
   7.2. Success Factors ............................................................................................................................... 22
   7.3. Benefits for the Region ..................................................................................................................... 23
8. Conclusion ............................................................................................................................................... 24

QA process

<table>
<thead>
<tr>
<th>VERSION</th>
<th>SUBMISSION DATE</th>
<th>COMMENTS ON PROCESS</th>
<th>AUTHOR / INSTITUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>JUNE.2020</td>
<td>Request for QA by external expert</td>
<td>Blaž Zupan/Matjaž Mikoš</td>
</tr>
<tr>
<td>1</td>
<td>JUNE.2020</td>
<td>QA Feedback by SAE</td>
<td>Stane Pejovnik/SAE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Janko Burgar/SAE</td>
</tr>
<tr>
<td>2</td>
<td>JUNE.2020</td>
<td>Request for QA by external expert</td>
<td>Blaž Zupan/Matjaž Mikoš</td>
</tr>
<tr>
<td>2</td>
<td>JUNE.2020</td>
<td>QA Feedback by SAE</td>
<td>Stane Pejovnik/SAE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Janko Burgar/SAE</td>
</tr>
<tr>
<td>3</td>
<td>12.7.2020</td>
<td>Send to DLR for Submission</td>
<td>Janko Burgar/SAE</td>
</tr>
</tbody>
</table>
1. Introduction

The South East European International Institute for Sustainable Technologies (SEEIIST, henceforth, the Institute) was proposed in late 2016 by Prof. Herwig Schopper, a former Director-General of CERN and initiator of the international SESAME project in Jordan. In 2017, the proposal received support from eight countries, including the Republic of Albania, Bosnia and Herzegovina, Republic of Bulgaria, Kosovo, Republic of North Macedonia, Montenegro, Republic of Serbia and Republic of Slovenia, and the initiative was transformed into a regional project after signing the Declaration of Intent in October 2017 at CERN.

In 2018, at the second SEEIIST Scientific Committee meeting, the parties decided the project will support Hadron Cancer Therapy and Biomedical Research with Protons and Heavy Ions as the option for the Institute. Hadron therapy is an advanced radiotherapy technique that uses proton or ion beams to deliver precise treatment of tumors, sparing the surrounding healthy tissues from unwanted radiation. This technique’s intrinsic precision makes it particularly suitable for treating tumors in children or close to organs at risk. Hadron cancer therapy is an alternative to conventional radiation therapy, where high energy photons are produced by accelerated electrons and then delivered to the patient to destroy tumor cells. Radiation oncologists irradiate the tumor target using crossing beams from many angles while trying to spare the surrounding healthy tissues. Inevitably, in conventional radiation therapy, some radiation dose is always deposited in healthy tissues. In contrast to photon beams, particle beams exhibit a Bragg peak in energy loss through the body, delivering their maximum radiation dose at or near the tumor and minimizing damage to surrounding healthy tissues.

A center for hadron therapy in south-east Europe has yet to emerge and is subject to activities of the upcoming Institute. In Europe, the interest in hadron therapy increased, and the first dual ion (carbon and protons) clinical facility in Heidelberg, Germany, started treating patients at the end of 2009. Three more such facilities are now in operation: CNAO in Pavia, MIT in Marburg, and MedAustron in Wiener Neustadt are treating patients.1 Globally there is considerable momentum in particle therapy, especially treatment with protons. By 2020, there will be almost 100 centers around the world, with over 30 of these in Europe.

In proton therapy, medical personnel uses a particle accelerator to target a tumor with a beam of protons. The proton accelerator must produce a higher energy beam and typically produce protons with energies in the range of 70 to 250 MeV. The proton therapy center’s core equipment consists of an injector (ion source), synchrotron, and irradiation chambers.2 Irradiation chambers can focus the proton beams anywhere by rotating the gantries 360 degrees. It is possible to avoid areas we do not wish to irradiate and project the beams at the back or the sides of the body, or any other angle desired. The radiation equipment is software-controlled and uses a range of algorithms and technologies to determine the positioning of the irradiation chambers and control all aspects of the treatment.

The Institute will aim to establish a cancer treatment facility, but it will also complement the clinical aspect of the work with research in new radiation technologies. In particular, the Institute will focus on patient treatment with carbon ions and protons for about half of the daytime and first in two and later in four treatment rooms, treating about 250 to, later, 500 patients per year. The established treatment should cover a significant fraction of the yearly number of Southeast Europe patients having tumors of the highest priority for

---

1 https://enlight.web.cern.ch/what-is-hadron-therapy
2 https://www.nptc.city.nagoya.jp/e_proton/e_equipment.html
carbon and proton irradiations. For the remaining time, the radiation equipment will be devoted to research work, focusing on:

1. *in vitro* radiobiology experiments, to better understand the fundamental mechanisms of radiosensitivity and radioresistance,

2. animal studies for *in vivo* determination of the efficacy of carbon and other ions in the treatment of human radioresistant and radiosensitive tumors, and healthy tissues’ effects,

3. medical physics measurements and development of novel radiation detectors and optimized treatment planning systems.

**We foresee that all activities of the Institute will be data driven.** Digitalization and use technologies from data science and artificial intelligence are necessary for any application of modern cancer treatment. They are the foundation of state-of-the-art research in any branch of science. The Institute should thus consider digitalization and the use of the latest computer science technologies from the start. Digitalization will be managed and supported by the Institute’s Digital Hub, an entity that will care for all aspects of data gathering, curation, management, and data-driven support of clinical and research part of the Institute.

The main aim of this document is to define the potential within the Institute and its digital agenda that should explore the regional potential. We will examine the necessary capacities and opportunities for introducing a Digital Hub and study add-on benefits for the region. The Digital Hub might become a complementary research and development area supporting the scientific and clinical part of the Institute. The Institute’s digitalization will be driven by the needs of data collection, data-based support of daily clinical operations, data science support of the research, and collection of approaches founded in artificial intelligence to push on the frontiers of science and medical research.
2. Scope & Areas of Research and Innovation in the Digital Hub

As one of the Institute's central units, the Digital Hub will establish and manage the data infrastructure, develop and maintain software applications, and perform research in new algorithms and mathematical procedures to improve and support clinical and research. It will establish technologies that will specifically address the operations of the Institute. Here, we aim to precisely define the scope of the infrastructure that will be managed by the Digital Hub and the set of research and innovation activities that will take place. Digital Hub will be a center for support and a center for research in data science applied to hadron radiation technologies.

2.1. The Scope

Digital Hub will direct and manage all facets of the digitalization of the Institute. We here layout the core components of the digitalization, as this will help us to reason on Digital Hub’s equipment, staff, and processes. The scope of the work of the Digital Hub should include the following items:

**Hardware infrastructure.** The Data Hub will rely on the following core set of hardware:

- **Servers.** We foresee that the Institute will collect both smaller datasets and possibly extensive collections of data related to the measurements of all aspects of radiation. Extensive data collections will fall within the big data category, where raw data storage takes substantial disk space and requires specialized software for the analysis. Servers will thus include file servers and computational servers. One type of stored data will possibly be images, and for their study, the Data Hub will use artificial intelligence techniques that require processing by graphical processing units (GPUs). Besides standard multi-core processing servers, the use of AI will require GPU servers. We propose that the Institute builds its hardware infrastructure gradually and conservatively, and make use of all available hardware from regional partners to reuse available resources and economize with the budget.

- **Sensors.** Radiation therapy equipment involves complex components that can be monitored and controlled. Especially for the research, data collection will require specialized hardware equipment in the form of internet-of-things (IoT), that is, interconnected sensors and actuators that communicate with the cloud for data storage and exchange. The data from the IoT cloud is then preprocessed and analyzed at the data analysis layer. The IoT may become an essential part of the Digital Hub and require specialized expertise for its set-up and management.

- **Networks.** Hardware infrastructure is tight together through a network. Though this may seem obvious, construction and management of the system require specific skills. They must be well-supported within the Digital Hub, both in terms of the expertise and sufficient redundancy in the staff to avoid any significant bottlenecks. Keeping the network alive is essential for any data-driven enterprise.

**Data infrastructure.** The Data Hub will collect and manage all kinds of data that will support clinical operations and research of the Institute. The essential items of this infrastructure include:

- **Databases and database management.** The Data Hub should be capable of establishing a set of databases to support in-house data collection. Creating and maintaining the databases entails the employment of database developers and managers, with profound knowledge of database management software. An
alternative option would be to rely on third parties for this task. As data will be Data Hub core resource, we would strongly advise against giving it in the hands of outside collaborators for lower reliability, more extended time response, and difficulties in understanding the core processes at the Institute.

- **Data security and safety.** To support the clinics, the Institute will deal with personal data prone to the General Data Protection Regulation. The Institute will also collect the research data that will be private and should be kept safe. Besides safety, both types of data require backup procedures. Data security and safety are becoming a complex issue and need specialized staff with experience complementary to the development and maintenance of databases.

- **Application program interfaces.** Databases need to implement interfaces to data access to pass the data to layers of software applications and analysis infrastructure. Data Hub should plan these interfaces ahead of time, providing a uniform framework across the data infrastructure and taking care of their implementation, testing, and documentation.

**Business intelligence and decision support layer.** We see the Institute as a data-driven organization. The clinical part of the Institute should become a model for patient management and treatment planning, and as a side effect, provide any required data for clinical and treatment research. In part, the experience for developing and maintaining such technology should transfer to research activities as well, but primarily the role of this layer will support the clinics.

- **Dashboards and data monitoring.** Dashboard technology is crucial to assist non-technical users in their daily activities, like physicians reviewing their patients’ data or radiation engineer planning a radiation session. Which sections of the clinical data will be accessible in this way is determined by domain experts. It is essential that the know-how of constructing simple user interfaces to the collected clinical data is crucial and should reside within the Digital Hub.

- **Decision support.** Creating algorithms and procedures to support physicians in treatment planning and prognosis is an area of research within artificial intelligence in medicine. Today, standard procedures are collecting the clinical data and then using machine learning to develop models for decision support. The principal challenges in this field include assessment of reliability and implementation of effective decision support systems that can seamlessly complement the daily decision making. Digital Hub will have data at its disposal and can, in collaboration with domain experts, cherry-pick the applications that would require and benefit from machine-based decision support.

**Data science layer.** Digital Hub will collect a possibly massive amount of data and will hence provide services and expertise in data analysis. Data analytics involves the following key components:

- **Big data infrastructure and high-performance computing.** The data from radiation equipment, which includes the data from daily operations and the experimental data, and the data to monitor the Institute as a Green Hub, will be substantial in volume and scope. The Digital Hub will need to establish big data software infrastructure to process and analyze this data. For some aspects of the analysis, it will require high-performance computing and corresponding analysis procedures to support parallel querying, summarization, and abstraction of the raw data.

- **Data science infrastructure.** With data science infrastructure, we here refer to software for data analysis, particularly the software to support statistics and machine learning. The Digital Hub will need to manage this software. More importantly, it will need to include data scientists that will use the tools and support the research using the available data. With this infrastructure, we specifically address data visualiza-

---

3 https://gdpr-info.eu
tion and data exploration procedure, which should be accessible to researchers and staff of the Institute who are not necessarily computer science experts.

- **Deep learning layer.** The data on images and sensor data will require processing and modeling with deep learning. It is necessary that the Digital Hub addresses this part of artificial intelligence and maintains the software infrastructure and in-house expertise.

Notice that the items that we have listed above and are within the scope of the work addressed within the Digital Hub support various types of users (Figure 1). The Digital Hub will need to provide support for all ends of the user spectrum, starting from the developers of software services to the end-users, including physicians and patients. The range of required expertise for such support is broad. To be fully equipped and staffed, the Digital Hub will need to rely on a vast array of information technology solutions and software technologies and, more importantly, on its employees and collaborators’ excellent staffing. See section on staff for a discussion of this later issue.

![Figure 1: Different types of users and their relative size addressed by the activities of the Digital Hub. The largest group of users are non-technical users who require applications with a sufficiently simple user interface. Examples are decision support systems for physicians or even phone apps to support treated patients. On the other side of the spectrum are technical users that receive support from the Digital Hub in the form of access to the infrastructure, set of documented data interfaces, or access to in-house expertise.](image)

### 2.2. Research and Innovation

The primary function of the Digital Hub is to provide data-driven support for all operations of the Institute. Complementary, though, the Digital Hub will also perform research in data science and decision support and will have a huge opportunity to innovate in this new area of medical applications and research in radiation. In particular, we

foresee the utility of all aspects of machine learning and innovations based on artificial intelligence applications, particularly in:

- the machine learning techniques and data science applied to proton therapy current problems, where there is opportunity to innovate because of the young age of the field,
- machine learning and large-scale analysis of sensor data (IoT data), where the opportunities lie in the intersection of sensorics, signal processing, and machine learning,
- data fusion, which should address in-house integration of data from different data sources,
- image analytics applications, which stem from the potential to collect a vast number of labeled images pertinent to proton therapy,
- data science techniques in decision support will stem from opportunities due to the gathered collection of patient data.
3. Challenges

Principal challenges in establishing Digital Hub are:

- **Company culture.** It is vital that Digital Hub becomes an essential part of the Institute right at the beginning of its incorporation, and that the Institute involves the staff from the Digital Hub in every aspect of planning and management of the organization. It is much easier to set up a data-driven company right from the start than to change the already established classical company into a data-driven one (see Figure 2).

- **Management structure.** The director of Digital Hub should be a board member of the Institute. Data infrastructure and digitalization are today probably an essential part of every organization, and the ability to participate and influence crucial decisions based on the understanding of the data processes is vital.

- **Staffing.** Data scientists, software engineers, data safety specialists, and other profiles from computer science are today hard to find on the job market. The Institute should consider that experienced computer scientists, and especially data scientists are today among the best-paid jobs. Hiring the staff for the Data Hub will require good advertising, PR, and job benefits.

- **A wide range of expertise needed.** The employees of the Digital Hub will have to cover a wide range of knowledge. Staffing should consider and focus on job candidates that can learn and adapt fast, are already experienced in at least one of the expertise required by the Digital Hub, and are good team members.

- **Interaction with domain experts.** The collaboration between the Digital Hub and other departments of the Institute will be crucial to the endeavor’s success. These collaborations have to be set right from the start of the operations, or, even better, at the planning stage of the Institute.
Figure 2: Bottlenecks to AI adoption reflect the most significant obstacles in introducing advanced data science technologies in enterprises.5

- **Establishment of reliable and uniform data infrastructure.** Data infrastructures often suffer from the entropy of various software solutions, often inflicted by the third-party developers. It is possible to circumvent this challenge by clear goals and strong leadership of the Digital Hub and the establishment of solid engineering guidelines.

- **Data protection and privacy.** The Digital Hub needs to prioritize data protection and privacy issue by careful staffing. There should be at least one specialist employed that covers this area directly, and the Digital Hub should educate all of their development engineers to be knowledgeable in this field.

- **Choice of software tools.** We would recommend Digital Hub to use open-source tools. Notice that open-source software tools are already well established within computer science. For example, the most used tools in artificial intelligence are in open source (Figure 3). It would be unwise and unnecessarily expensive to use commercial solutions where open source software leads in quality and provision of modern approaches.

- **Lack or excess of hardware equipment.** The Digital Hub should be aware that the highest incurring costs in data support for the companies are in the staff salaries, and not in the computer equipment. Especially research institutions, like the Institute, could benefit from collaborations with academia that already has access to suitable hardware that is often underused and could be, at least in the beginning stages of the Institute, borrowed or leased. The Institute should plan its purchase of servers carefully and avoid purchasing computational equipment if used only occasionally.
There is a set of opportunities that the Digital Hub can ride on and from which the Institute can benefit substantially. These include the following:

- **The field is young: the right timing.** The area of proton beam therapy is just emerging\(^6\) and there are plenty of research and confirmation opportunities. Any data-driven center will have a huge advantage that could lead to fast improvements, optimizations, and research opportunities, including opportunities to publish.

- **The Institute is data-oriented.** Establishing the data infrastructure right at the start and setting the Institute culture in data-driven direction will be a huge gain and an advantage towards older, more complex, and already established organizations that are yet to adopt data-driven thinking.

- **Collaboration with partner institutions.** The Institute will be set up through a network of collaborating countries and their institutions. This network could provide for successful staffing, scientific interchange, interdisciplinary collaboration, and market opportunities.

- **Digital Hub is established at the very start of the Institute.** The fact that Digital Hub will be created right at the beginning of the Institute's operations could be very beneficial. Data infrastructure will be an essential part of the Institute. Early set-up of the Digital Hub will guarantee that the infrastructure will emerge early and will be optimized for quality and service.

- **Digital Hub is a core facility and an essential part of the Institute.** The Digital Hub should dictate the composition of the computational infrastructure of the Institute, define the data flow within the institution, and it establishes "the glue" that integrates various departments and teams. Exposing the Digital Hub as a core facility presents an opportunity for the Institute and the advantage towards similar institutions that push information infrastructure services lower in the organization hierarchy.

- **The director of the Digital Hub is a member of a Management Board of the Institute.** This item is conceptually related to the previous two. It exposes the organizational opportunity that the Institute has in setting digitalization and data science as a top priority in the Institute.

- **Early adoption of AI, data science, and machine learning in proton therapy planning.** With accumulated data, possibly from existing proton therapy centers, we can mine the libraries of previous patients treated with proton therapy to extract knowledge. This knowledge could then be distributed amongst the growing proton network to help bridge learning curves and provide a quality assurance framework.\(^7\) The utility of such data could, for instance, be used to develop models for online proton range and dose verification based on measurements of proton-induced positron emitters,\(^8\) or for prediction of monitor units for a compact proton machine.\(^9\) Early adoption of machine learning techniques could provide models that

---


\(^7\) [https://www.varian.com/resources-support/blogs/proton-therapy-insights/unleashing-power-proton-therapy-machine-learning](https://www.varian.com/resources-support/blogs/proton-therapy-insights/unleashing-power-proton-therapy-machine-learning)


increase the quality of service. Their application of machine learning and data science on the Institute's emerging data can lead to a better understanding of proton therapy's clinical effects and understand the factors that could improve the quality of the treatments. Currently, the use of AI and machine learning in proton therapy planning is scarce. Their early adoption presents both opportunities for clinical research and a chance to improve the quality of daily operations in the Institute substantially.

- **Early adoption of deep learning.** Deep learning methods have dramatically improved state-of-the-art speech recognition, visual object recognition, object detection, and many other domains such as drug discovery and genomics. The Institute can adopt deep learning for processing and modeling of its image database, its database of CT scans, and its emerging database of sensory data. The applications of deep learning could span across both the clinical and research parts of the Institute. Early provision of the data infrastructure within the Institute is a critical and reliable backbone for the adoption of deep learning. An early introduction of deep learning to data analysis could lead to early international recognition of the Institute. Namely, while deep learning is today the most active research area within artificial intelligence, its applications to proton therapy data are scarce. For instance, a recent example uses deep learning for automatic contour propagation for online adaptive intensity-modulated proton therapy of prostate cancer.

- **Institute-wide data integration.** The Digital Hub will have an opportunity to organize and manage the entire data infrastructure for the Institute. This presents data integration opportunity, that is, developing predictive models and data presentation dashboards that summarize and join information from different data sources. Especially in medicine, no single data type can capture the complexity of all the factors relevant to understanding a phenomenon such as a disease. Data integration and procedures for data fusion can substantially raise the accuracy of the models and provide a more in-depth insight into the interaction of factors that govern phenomena under the study.

- **Funding opportunities for big data, data science and interdisciplinary applications.** Research in big data and data science is well supported by grant opportunities, especially when combined with a new application domain. A current funding niche is explainable AI, an emergent technology that supports both inferences of predictive models with high accuracy and the ability to provide insight into models and explain otherwise potentially complex domains. The principal funding source for such activities is EU grants (H2020 and alike), but these should extend to local, state-specific grants. Within Europe, each of the countries is now preparing or has already prepared a strategy for adoption of the artificial intelligence, and with it has been proposing a schema to found activities to locally propel research and applications of this technology. The Institute, wherever established, should search for both local and EU-based support, and gain through links with CERN and network of proton therapy centers in Europe.

---

5. Principal Competencies of the Digital Hub

The Digital Hub should cover a relatively wide range of competencies. These stem from items we have defined in the scope of the activities within the Digital Hub (see Section 2). Even considering data science alone, the set of competencies defined by the Association for Computing Machinery, the principal international society of computer scientists, is substantial (Figure 4) and ranges from those from mathematics and statistics, computer science, and competencies in application fields.\(^\textit{14}\) The essential competencies the Digital Hub would need to include and nurture are:

- **High-performance computing infrastructure** includes setting up and managing computational and storage servers and setting up the software technologies for large-scale data processing such as Apache Hadoop, Spark, Kafka, and similar.

- **Computer networks** create an operable and interlinked data environment and seamlessly integrate all various technologies of the Digital Hub.

- **Internet of things and sensorics**, to set up the measurement infrastructure, connect it to storage clouds, and provide programmatic access to data analysis.

- **Databases**, including the development and maintenance of database infrastructure and particular databases. The development includes the definition of database architecture, the implementation of the database on a dedicated server, definition of user interfaces to access and manage the data, and provision of the data to higher-level applications, such as those for data visualization, business intelligence, and deep learning.

- **Computer safety and privacy** are essential competencies, especially when dealing with patients’ data from the clinical part of the Institute. Computer safety is also an issue of keeping the servers alive and protecting them from external computer attacks.

- **Data engineering** focuses on the practical application of data collection and analysis and prepares the raw data for analysis, visualization, and machine learning. Most often, data engineering deals with the construction of features that characterize and profile the object of interest. Data engineering has a profound impact on the quality of data analysis and modeling.

- **Statistics**. While competency in statistics may be covered by other departments as well, Digital Hub would benefit from Institute-wide support of statistical analysis to complete its portfolio and control the entire data processing pipeline.

- **Data science** is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from many structural and unstructured data. Support for all facets of data science is essential, and the data hub should provide it both in terms of equipment (servers), supporting staff, research staff, and provision of software frameworks.

\(^\textit{14}\) ACM Data Science Task Force (2019) Computing Competencies for Undergraduate Data Science Curricula, initial draft, Association for Computing Machinery.
- **Artificial intelligence and deep learning**, where the introduction of these technologies will be essential to push the Institute’s scientific boundaries and make it competitive and visible within the network of proton therapy centers. We foresee that these will become core technologies in data management and modeling to increase the service quality and understand the underlying principles.

- **Decision support systems**, mainly when applied to the clinical part of the Institute, is a core competency to provide support to clinicians and potentially to patients. Interdisciplinarity, teamwork, problem understanding, and construction of the right user interfaces are essential to delivering decision support systems.

- **Machine learning**, where we here refer it to technologies that are also outside deep learning and include feature selection, modeling of smaller data sets with, say, regularized logistic regression, ability to computationally end experimentally validate the resulting models, reliability estimation, and interpretation and explanation of models and patterns found in the data.

- **Data mining and data visualization**, where we here primarily refer to the construction of explorative data analysis interfaces for the end-users. The democratization of access and data analysis procedures is crucial to the acceptance and adoption of digitalization within any institution, and the Institute should embrace it throughout the entire enterprise and in all of its activities. The democratization of data science is possible with the right tools (e.g., for image analytics\(^1\)) which should be nourished in the Digital Hub.

- **Dashboard technologies and application development.** The data in the summarized form will be often accessed through the web interfaces and within the dashboards that should be constructed in-house because of understanding of the problem domain and ability to adapt to changes and user requests.

- **Communication and teamwork** is a soft skill that is essential to connect employees in the Digital Hub with the end-users. Nourishing these skills should be planned and systemized.

- **Ethical issues.** The Digital Hub will deal with the data that will stem from the clinical part of the Institute and its research operations. The ethical issues will be raised wherever the data from patients and human studies are gathered, stored, and processed, so the Digital Hub must be aware and address any ethical issues related to the management of this data.

---

Figure 4: Data science competencies as defined in the recent report by Association for Computing Machinery.14
6. Staff

The Digital Hub should employ and engage external collaborators to cover the competencies listed in the previous section. We here distinguish between the core staff that is essential for the Hub, the desired supporting staff that should be employed within the house to achieve the full potential, and the external collaborators, which may include experts from academia, SMEs, and students.

6.1. Core Staff

- director of the Digit Hub, with experience in management, operations, computer applications in bio-medicine and health care, and knowledge about high-performance computing, data science, and artificial intelligence,
- senior computer technician to set-up and manage the servers and network,
- database developer, to design the databases' architecture and implement and maintain the databases in the Institute,
- senior data scientist to cherry-pick, plan, design, and apply data science support and systems and identify data applications with the highest potential and chance of success,
- junior data scientist to assist in implementing data science and collaborate in practical implementations of data science across various problem domains. Junior data scientist can also help the database developer,
- application developer, with excellent knowledge of web-programming and development of dashboards,
- data visualization and web application designer, wherein inclusion of this staff member we follow good practices of several genome centers (for instance, New York Genome Center), where they realized that excellent communication of results is essential for any data-driven operation.

6.2. Supporting Staff

- security and data privacy expert,
- junior database developer,
- junior application developer,
- senior high-performance computing engineer,
- senior project manager and grant proposal writer,
- statistician,
- internet-of-things expert,
- deep learning and machine learning expert,
- data engineer.
6.3. External Collaborators

- expert in data ethics,
- database developers,
- data scientists,
- application developers,
- data fusion and data integration experts.
7. Regional Considerations

South-east Europe currently does not have an established center in proton cancer therapy. Still, each participating country has shown capabilities in establishing and managing data centers and directing data science operations linked to various industries and research. In particular, some of the participating countries are among world-leaders in technologies that we have mentioned in this report. In particular, researchers in some countries led research in artificial intelligence, organized perhaps Europe’s largest data science gatherings, injected data science in current industries, and educated world-leading computer programmers. It is beyond the scope of this report to, in detail, analyze the capabilities specific to each of the countries. Instead, we stick to the extent of the previous chapters and state what we expect for these capabilities to include and when choosing the placement of the Institute, what factors may affect the success of its operations.

7.1. Present Capabilities

When considering the Institute’s placement in establishing successful Digital Hub, decision-makers should consider the following capabilities of the local environment:

- vicinity of a major educational institution with demonstrated strength in computer science,
- available local expertise in high-performance computing and proximity of centers with high-performance servers. Here, the vicinity of such center is essential not because of the equipment but rather due to the availability of supporting staff,
- local experience in data science, especially in the academic environment, but also in the local SMEs that could participate both in setting up of the Digital Hub’s infrastructure and in joint projects,
- local experience and demonstrated capabilities in artificial intelligence and machine learning, including in image analytics. This primarily refers to the experience gathered at neighboring academic institutions,
- local experience in biomedical informatics, where we here refer to experience in communication with domain experts, know-how in cherry-picking biomedical applications, and knowledge of setting-up useful systems for decision support with the appropriate user interfaces.

We understand that the Institute and its Digital Hub will not limit its hiring from the local environment. Yet, collaboration with local institutions may be crucial in laying the foundations of Digital Hub’s activities and provide a spark for immediate activation of research and development within the Hub. These foundations will be essential when looking for joint international collaborations and could provide for initial expertise and references.

Countries in South East Europe can build on foundation of excellence in studies in math and statistics, education of engineers, legacy and emergence of IT industries, and pockets of AI research with high density of machine learning experts and data scientists. In a recent article, for instance, Financial Times reports on Belgrade’s fast-growing tech start-ups and growth in the number of ICT specialists in Serbia and in eastern European countries (Figure 5).
Figure 5. Serbia’s tech industry is growing fast, along with some other eastern European countries.16

There are numerous examples of emerging centers of excellences, small enterprises, and large-scale collaborations that are emerging in this part Europe. Here, we only mention a few examples, as a thorough review of capabilities of each of the participating countries is beyond the scope of this report. Yet, example we list here shows that the partner countries are capable in competing and collaborating within European research space and markets, and need challenges such as SEEIIST to boost their advancement. Following are therefore few examples of success stories:

- **GATE** ([https://www.gate-coe.eu](https://www.gate-coe.eu)), a Big Data and AI Centre of Excellence coordinated by Sofia University "St. Kliment Ohridski", Bulgaria, aims to become as a globally competitive ecosystem for conducting basic and applied research and innovation in the Big Data areas with high social impact. Its operations have started in late 2019. The project is coming to life as a joint initiative between Sofia University, the Swedish Chalmers University of Technology and Chalmers Industrial Technologies. With over 50 similar setups in Western Europe, GATE is the first of its kind on the Balkans and has managed to secure €15 million in funding from the EU’s Horizon 2020 program.

- Institute of Contemporary Science, a non-government institution in Belgrade, Serbia, is organizing some of the largest and best conferences in south-eastern Europe, including Data Science Europe 2020 in Belgrade, and Data Science Conference Croatia. (See [https://www.datasciconference.com](https://www.datasciconference.com) and [https://croatia.datasciconference.com](https://croatia.datasciconference.com)).

- Seven Bridges ([https://www.sevenbridges.com](https://www.sevenbridges.com)), a private international company that was founded and runs from Belgrade, Serbia, leads genomic and healthcare data analytics and is known for pioneering its solutions based on graph databases, perhaps currently the most modern data engineering and science approach.

- The Event Registry ([http://eventregistry.org](http://eventregistry.org)) is an example of an entirely Slovenian project and product developed with the aid of European funding. The tool is a system that analyzes more than 100,000 text

---

16 Hopkins V (2020) Belgrade’s fast-growing tech start-ups show Serbia’s ‘hunger for success’. Financial Times, June 17, [https://www.ft.com/content/03f995f0-9c12-11ea-871b-8deb99a20c6e](https://www.ft.com/content/03f995f0-9c12-11ea-871b-8deb99a20c6e).
news sources in various languages to track how news spreads from one location to the rest of the globe, including story development and the context in which the news is mentioned.17

- Orange (https://orange.biolab.si) is a popular open-source data science framework that has been used around the globe for data analytics as well as teaching of machine learning. It features innovative visual programming interface and interactive data visualization, and is being developed by University of Ljubljana, Slovenia.17

- University of Belgrade ranked within top 400 universities in the world includes an internationally-accredited Faculty of Organizational Sciences (http://www.fon.bg.ac.rs/eng) with excellent research in decision support and machine learning, and with excellent connections with flourishing AI community in Belgrade, Serbia.

These are only a few examples where centers in the countries participating in SEEIIST established projects and programs with strong international presence that demonstrate their technological and staff readiness for projects such as the one planned within SEEIIST. While selecting which local environment would fit best for supporting the proton therapy with IT and AI is beyond the scope of this document, we here pointed out at criteria and factors on which these environments can be judged. We also stress that most, if not all of the participating countries have centers with local capabilities both in personnel and equipment for such a support.

7.2. Success Factors

The Digital Hub’s success will depend on how well it addresses the challenges and to which width and depth it can cover the capabilities we have outlined before. From experience from establishing similar centers and project, we propose that the following factors will provide keys to success:

- The Digital Hub is recognized as the core facility within the Institute.

- The director of the Digital Hub is placed within the core team that runs the Institute, and is, for example, on par with the director of the clinical and director of the research department.

- Digital Hub’s director spends fifty percent of the time in the first two years devoted to hiring the best possible staff.

- Initial Digital Hub’s projects are cherry-picked to maximize project success and visibility and provide a model for other upcoming projects.

- The staff of the Digital Hub has established excellent communication with the rest of the Institute, and the employees of the Institute are convinced that the Digital Hub contributes to their success in a major way.

- The Digital Hub has established itself both as a service provider and as a research unit. The outside impression is that the Hub can solve problems and invent new technologies to optimize and improve daily operations.

- Local government supports the Institute and, with it, the Digital Hub. The support includes the financial element, which will be essential in the early years of operations. The founders of the Digital Hub should make a convincing argument about potential benefits to the region’s development.

These factors are closely related to a region that will establish the Institute. The region will be instrumental in sparking all the initial efforts, recognizing the support of the Digital Hub, and promoting it locally, within South East Europe, and internationally.

7.3. Benefits for the Region

Establishing the facility for clinical application and research of proton therapy may provide substantial benefits for the region. The Institute will be the first such center in the area that will benefit from early adoption, an excellent network of centers, and enthusiasm about new technology and implications for health care. The Digital Hub, on its own, will impact the region as well. The data-driven economy has lately been the enterprises’ driving mantra, with the digitalization of everything the leading transformation of businesses. The Digital Hub will

- gather and nurture expertise across the full board of digitalization,
- provide a case study of the introduction of data-driven operations within an upcoming institution from the ground-up. By demonstrating its success, create a model for such endeavor locally, to be followed by other enterprises and institutions,
- spark research in data science and artificial intelligence,
- join the Institute’s public relations efforts to communicate the benefits of new data technologies and awareness of how artificial intelligence can help advance medicine and health care,
- spark interest for the engineering studies, and through linking biomedicine and engineering, raise interest in the topics in these studies for the underrepresented female population,
- spark international collaboration of local institutions within EU programs through joint research and development,
- spark the interest in establishing new start-ups and spin-offs that could focus on aspects of digital support for protein therapy centers.
8. Conclusion

Establishing the Digital Hub within the South East European International Institute for Sustainable Technologies will in the most profound ways to determine the success of the Institute and its operations. The right execution of creating such a Hub, employing the most capable staff, and engaging them in fruitful collaboration with the rest of the Institute is crucial. It should be handled with care, in a planned, systematic way. The Digital Hub will face many challenges that are both social and technical nature, and starting with successful projects right from the start will be essential.

Responsible Organization for this deliverable:
- For SAE as external expert: Dr. Blaž Zuoan

Quality Assurance made by:
- SAE: Janko Burgar M.Sc, CGO & Project manager
- SAE: Dr. Stanislav Pejovnik, General secretary & Project member