

Newsletter #2

iTOBOS

Intelligent Total Body
Scanner for Early
Detection of Melanoma



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 965221.

2022



Contents

THE IMPORTANCE OF EARLY DETECTION IN THE LIGHT OF COVID-19

THE VALUE OF CLINICAL DATA

MACHINE LEARNING BASED APPROACHES FOR NON-CONTACT DERMOSCOPY

THE CHALLENGES OF INFUSING PRIVACY AND COMPLIANCE TECHNOLOGIES IN THE ITOBOS PROJECT

OUTSTANDING IMPACT OF ITOBOS PROJECT IN THE MEDIA

SOME PROJECT EVENTS

WORK PRESENTED

PUBLICATIONS

ITOBOS TEAM

LET'S STAY IN CONTACT!



The importance of early detection in the light of COVID-19

Melanoma has a poor prognosis with median survival of 6-9 months in the absence of timely diagnosis and treatment. Thanks to the introduction of novel, highly effective but also costly therapies (immune therapies with anti-PD1 and anti-CTLA4, and targeted therapies for BRAF mutant melanoma), overall survival has now considerably increased to over 50% at five years¹.

In this setting, **access to early diagnosis and surgical treatment, as well as continuous access to systemic therapies and monitoring, are critical for the survival of our patient population.** Under COVID-19, we have already witnessed a dramatic decrease in cancer diagnoses across all cancers, including melanoma². Early detection of melanoma was, even before the pandemic, a critical aspect in Eastern European countries such as Romania³. Additionally, the Romanian healthcare system is known as one of the most under-performing in Europe and has for several years in a row ranked last in the EHCI reports^{4,5}, thus we expected that

access to screening and melanoma care would become rapidly critical.

Under the COVID-19 pandemic and considering the large number of patient complaints we had in our online forums, and the worrying data reported by the Romanian Health Observatory⁷, the Romanian Melanoma Association and Melanoma Patient Network Europe (MPNE) initiated a study to capture how the lock-down affected the access to health care and how to potentially mitigate its negative effects. Our survey was completed by 108 respondents, members of the Romanian Melanoma Association, in the springtime of 2020. 62% were patients and 38% caregivers, who completed the questionnaire on behalf of the patient. We were not surprised to see that even in the early days, the COVID-19 pandemic has impaired the access to melanoma care for 40% of Romanian melanoma patients.

Comparable data reported in the Netherlands showed that 30% of cancer patients experienced access issues to

¹ Larkin, J. et al. Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *New Engl J Med* 381, 1535–1546 (2019).

² Dinmohamed, A. G. et al. Fewer cancer diagnoses during the COVID-19 epidemic in the Netherlands. *Lancet Oncol* 21, 750–751 (2020).

³ Forsea, A. M., Marmol, V. del, Vries, E. de, Bailey, E. E. & Geller, A. C. Melanoma incidence and mortality in Europe: new estimates, persistent disparities. *British Journal of Dermatology* 167, 1124–1130 (2012).

⁴ Euro Health Consumer Index 2018. <https://healthpowerhouse.com/media/EHCI-2018/EHCI-2018-report.pdf>.

⁵ Euro Health Consumer Index 2017. <https://healthpowerhouse.com/media/EHCI-2017/EHCI-2017-report.pdf>.

⁶ Euro Health Consumer Index 2016. <https://healthpowerhouse.com/media/EHCI-2016/EHCI-2016-report.pdf>.

⁷ Impactul Pandemiei COVID-19 asupra bolnavilor cronici_septembrie 2020-1.docx. https://health-observatory.ro/wp-content/uploads/2020/10/Raport_ORs-impact_pandemie_cronici_2020.pdf.

their oncological treatment or follow-up⁸, making our results consistent with the overall lower performance of the Romanian healthcare system⁴⁵⁶. What was the most worrying aspect is that more than 50% of patients delayed investigations such as dermoscopy, imaging and laboratory tests by their own initiative, while a third saw investigation delayed upon their physician's initiative. Early stage and stable patients were more likely to give up on their regular consultations, while those with melanoma progression were the most concerned. The fact that it was predominantly patients rather than physicians who delayed dermatological controls and follow-up visits by their own initiative means that even with sufficient protective measures in place, melanoma patients will need to be proactively informed about the risk of melanoma and reassured of their safety in order to seek care.

Education on secondary prevention and early detection, especially in the first years after diagnosis, seems to be

paramount. Our respondents were either members of our Romanian Melanoma Association and MPNE or followers, meaning more likely to be proactive, educated patients. Therefore, we consider these results as the “best possible scenario”, with even worse outcomes for those patients less able to understand their risk of melanoma or to navigate the Romanian healthcare system. This means that **patients from countries with already lower healthcare performance and less exposure to early detection measures will now be disproportionately affected by COVID-19 delays in cancer diagnoses**. To avoid the spiralling out of control of the melanoma situation in countries like Romania, we, therefore, see an urgent **need to implement innovative models for early detection both for melanoma patients and high-risk populations to at least dampen the future impact of COVID-19 on outcomes in melanoma**.

Our results were published in the Journal of Radiotherapy and Medical Oncology⁹.



⁸ de Joode, K. et al. Impact of the coronavirus disease 2019 pandemic on cancer treatment: the patients' perspective, *European Journal of Cancer*. Volume 136, 132 – 139.

⁹ Violeta Astratinei, Andreea Strambu-Dima, Bettina Ryll, Critical Issues in Melanoma

Follow-up and Treatment, Safety of Patients and Communication with Healthcare Professionals During Covid-19 Pandemic in Romania. *Journal of Medical and Radiation Oncology* 1, 15-33 (2021).

The value of clinical data

Over the last half century, clinicians and researchers have made enormous advances in detecting and treating melanoma. From the development of simple and widely usable clinical checklists to dermoscopy to 3D whole body imaging to artificial intelligence algorithms, and from operations to chemotherapy to today's targeted therapies and immunotherapies, new discoveries have helped detect melanomas early, when they are most easily treated, and improved the outlook for those with advanced melanoma. The common element in all these advances is clinical data: images, histopathology results and information from clinicians and patients alike who have generously donated their time, experience, and knowledge.

Clinical data can be collected piecemeal from case reports of unusual clinical cases, but to be most effective data needs to be collected in carefully designed cross-sectional or longitudinal studies. These studies must incorporate enough participants for meaningful statistical analysis, otherwise results may be unreliable. Standardised data collection protocols allow comparisons between participants and well-designed data management procedures ensure that the information can be used effectively, even when the original collectors leave the research group, or by other research groups with new research perspectives that could be tested on previously collected data.

A wide variety of data can be collected in clinical trials: personal and family medical history, number and size of moles, how easily the person sunburns,

outdoor work or leisure pursuits, and the fine details of a previous melanoma, including exact anatomic location, can all prove useful to researchers. Checking for links between the different types of data can also reveal surprising information. An unexpected finding from one of our studies is that Australian men with both many moles and two dysfunctional copies of the red hair gene have an enormous 23% lifetime risk of melanoma. Based on previous knowledge, we expected them to have a 9% risk. This knowledge helps us to prioritise these people for careful screening to detect their melanomas early. It also gives us insight into the mechanisms of melanoma development, which may ultimately help us to prevent melanoma altogether.

However, before any data can be collected, researchers must consider ethical clearance and informed participant consent. Universities and research hospitals provide Human Research Ethics Committees that include legal experts and volunteers not otherwise affiliated with research or the institution for a true outside perspective. They consider aims, proposed methods, and how the privacy of participants will be protected. It's not enough to just be curious about something – there must be at least a potential benefit to melanoma patients to participate in research. It's also important that participants understand what they are signing up for and be able to both consent and withdraw freely. Patient volunteers are our most precious asset, and their time, comfort and privacy are essential considerations to any clinical data collection.

Machine learning based approaches for non-contact dermoscopy

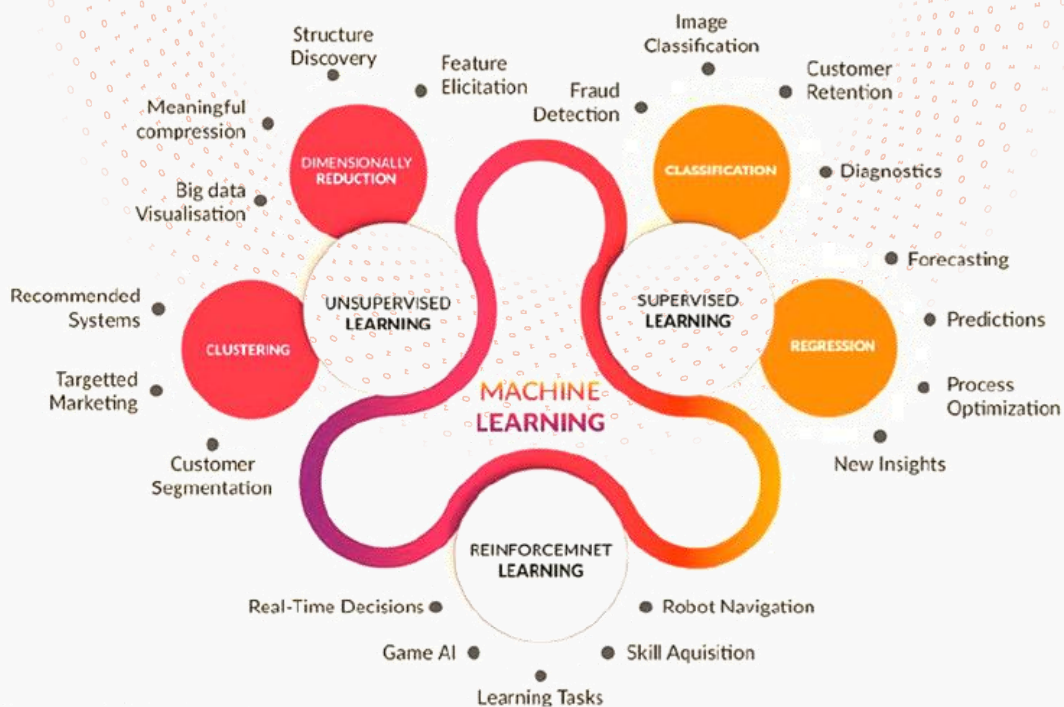
In iToBoS, machine learning/ artificial intelligence is key to combine all the design and make the system really a standout product.

Before going to nitty gritty of the project, we introduce some machine learning concepts here. Without the use of technical jargon, it should be easy enough for a technologist to understand and gain fundamental knowledge. According to dermatologists, melanoma can be detected with an absolute accuracy of 75-84% based on their visual screening of the skin images. Machine learning has a huge potential to automate diagnostic systems and it can give doctors a shot in the arm.

Detecting skin cancer using dermoscopy images can be achieved by improving deep learning capabilities. Classifying skin cancer lesions using specialized convolutional neural networks could help the medical community detect and potentially treat the disease in its early stages.

Overview

Nowadays, machine learning is a heavily researched and focused field around the globe. There is a variety of applications, such as in industry, medicine, and IT. Machine learning is an application of artificial intelligence (AI), which provides a system that allows learning and improving without explicit programming.



Supervised learning

An algorithm group that requires a pair of input-output data. For each pair of data, we take a sample of data used to make estimates and expected results as labels. Supervised is a word derived from the fact that human auditors must assign labels to data. During training, the model is fed with samples repeatedly. As for each sample, the model extracts the current status parameters and provides estimates. A difference between prediction tags is referred to as an error. Mistakes provide feedback on how to reduce errors in future forecasts. As a result, the model changes its parameters based on the algorithm from which it was derived. Supervised learning model tries to find for parameter values that enable it to achieve good results in historical data. They are then used to predict unknown data that is not part of the training set.

Supervised learning can solve two main problems of classification and regression. Classification categorizes input data instances. An example of use is to estimate if someone is sick, identifying false transactions, and identifying facial recognition. Regression estimates continuous numerical values for input data samples. Example of uses are house price valuation, food demand estimation or temperature estimation.

Un-supervised learning

A set of algorithms analyses unlabelled data without referring to known or marked results. No right answer exists for unsupervised learning. It is possible to explore unknown data structures and models using models based on this algorithm; a set of algorithms that attempts to extract conclusions from

unlabelled data (without referring to known or marked results). There is no right answer for unsupervised learning. Models based on this algorithm can be used to explore unknown data models and data structures. Unsupervised learning can solve two main problems of clustering and dimensionality reduction.

Clustering ensures that objects in different groups are not identical by assigning them to homogeneous groups (called clusters). The aim of clustering is to discover hidden objects by defining their structure. Dimensional reduction is very useful for high-dimensional data. This process aims to reduce the number of functions considered as part of a dimension that represents an object. With more features, data becomes scarce and analysis is corrupted by dimensionality. Using smaller datasets is easier.

Reinforcement Learning

Reinforcement Learning (RL) is a type of machine learning technique that uses feedback to help activists learn through actions and experiences. Both supervised and reinforced learning use mapping between input and output. However, unlike supervised learning, in which the feedback representative receives a set of corrective actions, reinforcing learning uses appreciation and punishment as a sign of positive and negative behaviour.

As opposed to unsupervised learning, reinforcement learning has different objectives. Compared with unsupervised learning, enhanced learning aims to find a model of action that maximizes the total cumulative reward of the model. Algorithms for Reinforcement Learning are numerous. The best action is determined by the current state of the results in Reinforcement Learning.

The challenges of infusing privacy and compliance technologies in the iToBoS project

In any data processing project that deals with personal information there is an inherent trade-off between safeguarding data subjects' privacy and yielding useful and accurate insights from the data. This problem is compounded on both fronts when dealing with healthcare projects: On one hand, health data is especially sensitive, and even defined under article 9 of GDPR as a special category of personal data to which additional restrictions apply. On the other hand, when making medical decisions that can have a “life or death” effect on people, you need to be able to make base those decisions on information and technology that is as accurate as possible.

Another dilemma is how to cooperate with the “open research data” paradigm being strongly promoted in the EU when it relates to health data, and deciding what data is safe to share outside the consortium. **The iToBoS project, for example, is interested in fostering the development of cognitive assistant algorithms, such as the one developed in the project, within the scientific community.** To this end, it plans to release some of the data collected in the project in the form of two open challenges for skin lesion analysis. However, for obvious reasons, any patient data shared publicly must be properly anonymized such that it can't be re-identified by any reasonable means, even with external knowledge. This means that merely removing direct identifiers such as names and IDs is not enough, and further measures to ensure proper anonymization of the released data must

be applied, for example k-anonymity or differential privacy.

When employing privacy-preserving methods on datasets, small amounts of noise and/or generalizations of the data are typically applied to prevent any record from being identifiable by better “blending in the crowd”. However, the smaller the dataset, the more difficult it is to create this blending, and the amount of noise or generalization required to reach that effect grows larger. Since the dataset in this project is relatively small (600 patients), this implies reduced utility of the published data.

Further compounding these issues is the fact that this project is considering analyzing more data sources than have been used in previous studies, including image data, clinical data, genomic data as well as family history. This makes it difficult to determine beforehand what parts of these data will be most useful for training the AI algorithms. Therefore, we cannot determine with certainty which fields can safely be removed/generalized or otherwise perturbed without harming the models' accuracy in a way that renders them useless for this type of application. We therefore plan to employ an iterative approach: initial versions of the AI models will be trained on (almost) all of the data collected during the prospective patient study. The importance or effect of each of the collected data points will be analyzed to guide the decision to possibly discard some of the features and not include them in the final training or challenges. For the remaining data which we do wish

to use, and include in the challenge data released publicly, we will attempt to create a tailored anonymization that takes into account the needs of the AI models and generalizes the data in a way that has a less harmful effect on those fields that are most important for the analysis.

The fact that the data has multiple modalities and formats also poses some challenges. The project will be employing the DICOM format to store, transfer and publish the image data in the project. The DICOM standard also includes many header fields to store metadata about each image, such as to which patient, study and series it belongs, what is the anatomical location of the image, along with additional demographic and clinical data which may be of relevance to the physician looking at the image and trying to make a diagnosis. In the case of iToBoS, some of these fields overlap with the data collected in the RedCap system and exported as csv files. In order to consistently mask or anonymize both data formats, the same masking techniques must be employed, with the same parameters and internal state, to yield coherent results. For example, it is necessary to mask the patient ID in the same way in both types of files so that all information belonging to a single patient can be cross-referenced. However, the standard tools used for de-identifying or anonymizing DICOM files do not currently contain the capability to mask tabular data, and vice versa.

Moreover, tabular data is not the only kind of data that may contain identifying information. Skin images may contain identifying marks such as birthmarks, tattoos or scars that, if included in the released data, may violate patients'

privacy. Image anonymization is a fairly new and yet unsolved research problem, especially in cases where very high resolution and level of detail is required, as in the case of dermoscopic images for skin cancer screening. In this project we decided to employ manual identification and selection of the identifying areas on the images by the technician operating the scanner, such that they can be removed from the image prior to being uploaded into the iToBoS system. One remaining challenge is how to ensure that the resulting image still looks smooth and natural enough to be processed by the system, including by AI models.

We also had to decide what to do with images of the facial area. On one hand, this part of the body is one of the areas with the most sun exposure and thus has high probability of sun damage and related disease, however these images are also potentially the most identifying. So as not to give up this important area entirely, we opted to use much smaller crops (1x1 cm) than those used for the rest of the body (6x4 cm).

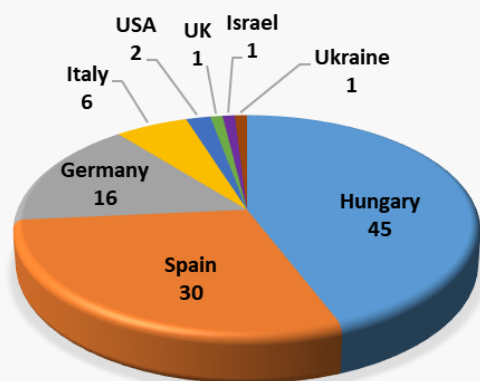
Finally, in a project such as this, that introduces a new AI-based decision-support technology, explainability of the models is of utmost importance to enable increased adoption of the tool by physicians and patients. Many methods for explaining AI decisions are based on revealing to the AI user specific samples (or characteristics of samples) from the training data used to train the model. This again poses a privacy risk to patients whose data was used in training the models. We are therefore considering ways in which auxiliary (non-private) datasets can be utilized in generating explanations for the AI decisions instead of the actual training data.

Outstanding impact of iToBoS project in the media

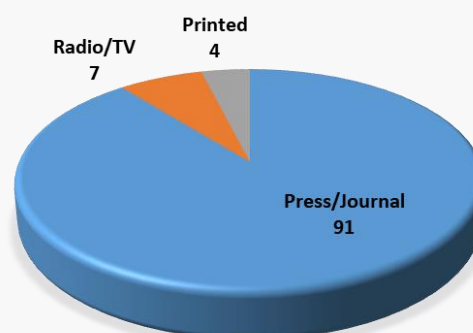
iToBoS project aims to provide new opportunities and added value to society in terms of **novel health solutions**, patient care, innovation, technical improvements and economic development. The interest of society has been high, as can be seen by the response and the intensity with which the media have dedicated to explaining the objectives, challenges, scope and opportunities of the project to wide and varied audiences around of the world, also considering different perspectives.

Since the beginning of the iToBoS project, a dissemination activity has been carried out with the aim of sharing the purpose and progress of the project, both from a general and technical approach, in order to reach the largest possible audience with the support of all the partners of the project. The international consortium has helped to reach and involve different national media, so iToBoS has been presented to their audience, considering the perspectives, profiles and specific interests of each one.

In the first year of the project, about **100 publications have presented iToBoS in media** in 8 countries: Hungary, Spain, Germany, Italy, United States, United Kingdom, Israel and Ukraine.



These disseminative articles have been produced and presented in different formats, both **thematic and general**, considering different formats: **digital press and journals, printed media and radio and TV channels**.



The publications present the project as an international and transversal initiative of the European Horizon 2020 programme. **They range from executive summaries to articles or interviews**. Technical data is also included and the social benefits inherent to the project are highlighted.

Through the presentation in **audio-visual media, the iToBoS leaders exposed in a more didactic and informative way the motivations for carrying out the project**, as well as the activities done, in progress or planned. Likewise, the work is presented in many **different languages spoken within the project** (Spanish, Catalan, German, Hungarian, Italian, Greek, French or Hebrew), apart from English as the vehicular language.

In the Media section on the iToBoS website, a summary list is presented with the most relevant articles related to the project published in international media. The activity is still alive, and more medias can echo the activities of the project and its proposal in the near future.

Some project events

In the second semester of the project, that covers from October 2021 to March 2022, iToBoS has organized and participated in different events, pointing out the following:

2nd General Assembly meeting

The iToBoS 2nd General Assembly (GA) meeting took place on 18th January 2022 throughout video conference system with the attendance of 49 participants belonging to the 19 project partner organizations.

In the GA meeting the project coordinator presented overall the status of the project deliverables, milestones and some administrative and financial issues. The WP leaders presented the current status of tasks and deliverables of their WPs, as well as the activities for the next 6 months.

“Ethical impact of the use of AI technologies for the detection of melanoma as addressed in iToBoS” Workshop.

The event was led by Trilateral Research Ltd (TRI) on 18th January 2022. The objective of the workshop was to engage project partners in the ethical risks associated with the development of the iToBoS tools and to work collaboratively to propose mitigation strategies. TRI leads WP2-Privacy, data protection, ethical and societal issues in iToBoS solutions.

2nd Project Management Board meeting

The iToBoS 2nd Project Management Board (PMB) meeting took place on

9/03/2022 throughout video conference system with the attendance of 17 attendees, including the Project Coordinator (PC), Project Manager (PM), Innovation and Exploitation Manager (IEM), Dissemination and Communication Manager (DCM), Data Manager (DM), Quality Assurance Manager (QAM) and WP leaders.

The updates in the tasks, the deliverables and the milestones of the project respect the last GA meeting were analysed in the meeting. The delays verified in some tasks and deliverables, some detected problems and the plan of contingency to solve them were reviewed as well.

Events participated

In addition, iToBoS representatives presented the project and shared experiences with a wide range of stakeholders, including relevant players from the information technologies, healthcare, research and innovation fields.

- Online, 4-8/10/2021. 26th European Symposium on Research in Computer Security.
- Online, 20-21/10/2021. XI Conferences on R+D+I & Health.
- Online, 09/02/2022. Green Project Expo.
- Barcelona, 28/02-08/03/2022. MWC Open Innovation Challenge 2022.
- Barcelona, 1/03/2022. 4 Years From Now Congress.

Work presented

During the second semester of the project, the following deliverables have been produced and submitted:

Deliverable submitted	Month	Leader	Diss. level
D13.1-H-Requirement No.1.	9	UQ	CO
D13.4-POPD-Requirement No.4.	9	SZTAKI	CO
D10.1-First study subject approvals package' for both the prospective clinical study and the clinical feasibility study (first release).	10	UQ	CO
D2.6-Privacy, data protection, social and ethical impact assessment report for iToBoS.	12	TRI	PU
D11.3-First dedicated training modules for Melanoma Patient Advocates.	12	MPNE	PU
D12.7-Exploitation and Business Plan.	12	RICOH	CO

Publications

During the second semester of the project, the following scientific publications have been developed in the iToBoS context.

- *"Explaining Machine Learning Models for Clinical Gait Analysis"*. 2021. Djordje Slijepcevic, Fabian Horst, Sebastian Lapuschkin, Brian Horsak, Anna-Maria Raberger, Andreas Kranzl, Wojciech Samek, Christian Breiteneder, Wolfgang Immanuel Schöllhorn, Matthias Zeppelzauer.
- *"PatClarC: Using pattern concept activation vectors for noise-robust model debugging"*. 2022. Frederik Pahde, Leander Weber, Christopher J. Anders, Wojciech Samek, Sebastian Lapuschkin.
- *"Explain and improve: LRP-inference fine-tuning for image captioning models"*. 2022. Jiamei Sun, Sebastian Lapuschkin, Wojciech Samek, Alexander Binder.
- *"Finding and removing Clever Hans: Using explanation methods to debug and improve deep models"*. 2022. Christopher J. Anders, Leander Weber, David Neumann, Wojciech Samek, Klaus-Robert Müller, Sebastian Lapuschkin.



In addition, different articles aimed at broader audiences have been developed and published on the project website, presenting the project from different scientific, medical, technological or innovation perspectives, considering the different profiles of all the project partners.

iToBoS team

The consortium with 19 partners organizations is led by the University of Girona (Spain). This international consortium brings together leading research/ academic institutions (5 research centres), industries (4 companies and 6 SMEs) and end-user entities (3 hospitals and 1 patients' NPO).



The University of Queensland has received funding from the Australia's NHMRC under grant number APP2007014.



Let's stay in contact!

iToBoS has deployed some **digital channels to keep in touch with you and bring you the latest news** about the project. They are also a way to receive your ideas and comments and learn more about your needs.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 965221.

The information provided reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.