Topic: Development of a platform for federated and privacy-preserving machine learning in support of drug discovery

All information regarding future IMI Call topics is indicative and subject to change. Final information about future IMI Calls will be communicated after approval by the IMI Governing Board.

Specific challenges to be addressed

Enabled by an ever-expanding arsenal of model systems, analysis methods, libraries of chemical compounds and other agents (like biologics), the amount of data generated during drug discovery programmes has never been greater, yet the biological complexity of many diseases still defies pharmaceutical treatment. Hand in hand with rising regulatory expectations, this growing complexity has inflated the research intensity and associated cost of the average discovery project. It is, therefore, imperative that the learnings from these data investments are maximised to enable efficient future research. This could be empowered by the big data analysis and machine learning approaches that are currently driving the digital transformation across all industries. These approaches not only rely on data generated specifically within a given project to learn from (as more established machine learning approaches tend to do), they also evaluate all other available data from different data sources and types for relevance to the question at hand. This extended approach will extract the maximum information present within the data, which in turns enables a gradual virtualisation of drug discovery processes and increases efficiency in bringing more and safer drug candidates towards clinical trials.

The success of the digital transformation in the pharmaceutical industry will thus highly depend on unlocking the maximal amount of data for the learning tasks at hand, and make these data amenable to the latest approaches in machine learning.To accomplish this, the following specific challenges need to be addressed:

1. Unlocking of proprietary and confidential data that is currently distributed across multiple data owners within the pharmaceutical sector without disclosure of the actual data and related assets themselves. In order to convince data owners to share their highly confidential and proprietary databases, which have been established over many years at considerable cost, the following conditions need to be fulfilled:.
   - **Privacy preservation** denotes the strict protection of the confidential and intellectual property (IP)-sensitive data and assets. In drug discovery, examples of IP-sensitive data and assets include the activity data of compounds in assays, the assay annotations, and predictive models derived from these data. In the strictest sense, privacy preservation implies that these data and assets never leave the control of the respective data owners.
   - **Federated machine learning** denotes here the distribution of the learning effort over physically separated partners. This goes beyond the currently more established concept of federated databases where the data are distributed, but not the data functionalisation (i.e. the learning from the data). It is key to enable owner control over data and other assets during learning.
2. Unlocking of data volumes from data sources or types that have hitherto remained untapped. In drug discovery, examples include image or transcriptional profiles or primary data points acquired in high throughput screens, all of which provide rich but hard-to-interpret biological annotation of chemical compounds.
3. Adapt recent advances in machine learning such as multi-task learning and deep learning for the above data expansion strategies.

Need and opportunity for public-private collaborative research

The digital transformation that is driven by ever more exhaustive data collection and exploitation, is disrupting the entire industrial landscape. Sectors and geographies that fail to embrace this transformation will find themselves challenged in their remit by newcomers with a strong footing in data sciences.

In this context, a collaboration among pharmaceutical partners, academia and knowledge partners from Small and Medium-sized Enterprises (SMEs) and other commercial organisations offers the perspective of doubling economies of scale in bringing better and safer drugs to patients. Firstly, it enables cost sharing and thereby bolsters the position of the European pharmaceutical industry in the global competition for data science and ICT resources. Secondly, it encourages data and method standardisation, thus expanding the volume of collective data that can fuel the big data revolution. Notably, these collective data should not be misinterpreted as a freely accessible and hence a fully precompetitive resource. Privacy-preserving approaches enable the reconciliation of collaborative investment with healthy within-sector competition.

The concepts of federated and privacy-preserving machine learning apply beyond the discovery remit, for instance in development and other clinical settings (like Real World Evidence settings). They even apply beyond the health setting. Indeed, by providing data owners the confidence that their data and the corresponding predictive models will remain private, the methodologies developed will encourage the formation of data and model consortia in various commercial (including non-pharmaceutical) and non-commercial contexts where data and knowledge ownership is at play. This creates opportunities for SMEs or other commercial partners that offer front-end or back-end services in the areas of software-as-a-service products in big-data analytics, clouded high-performance computing and privacy-preserving solutions. The public-private partnership proposed enables such partners to get exposed to, on the one hand, a strong application field with relevant use cases and clear ICT and security requirements, and on the other hand, academia and other knowledge partners with deep expertise in rapidly evolving science and technology fields

Scope

The topic aims for:

The delivery of a federated and privacy-preserving machine learning platform, initially validated on publicly accessible data, that is demonstrably safe enough (privacy-preserving in the face of legitimate and illegitimate (attempted) access and use) and scalable enough to be deployed to a significant representation of private data in the actual preclinical data warehouses of the participating major pharmaceutical companies in yearly evaluation runs. This effort will be mainly driven by the applicant consortium and enabled by the EFPIA partners.

The industry partners will subsequently drive the evaluation of the security, scalability and operational and predictive performance of the above platform on real industrial data (which is much more extensive than that in the public domain). As an indication of scale, the anticipated collective private compound and activity data sets from the industrial partners that will be used during the evaluations and that are to be accommodated comprehensively in each of the at least yearly runs, will include:

- at least 5 million chemical compounds annotated with dose-response quality activity data;
- at least 10 million chemical compounds annotated with some activity;
- at least 1 billion assay activity data points collected at single dose (low-complexity i.e. 1 to a few numerical values per compound, e.g. as from high-throughput screening);
- at least 100 million activity data points collected in dose response (over a range of doses, e.g. as from follow-up/secondary screening);
- several high-complexity activities collected at high-throughput (at least 100 thousand compounds in a standardised setting, e.g. high-resolution microscopy images or transcriptional profiles with 1000 readouts per well).

The above data are generated as part of the industry partners' normal drug discovery activities and, as such, are not generated in the scope of the project. Other than anonymized assay identifiers, the industry partner data will not include assay meta-information, such as specification of which drug target is tested. As a part of the effort, the industry partners will agree on protocols to standardise, format and normalise their private data for optimal interoperability and release software needed to do so.

The economic value of the platform lies in its ability to learn to predict the activity of chemical compounds in documented assays from descriptors of their chemical structure in the absence of meta-information such as the

drug target of these assays. For training the predictive models, the platform will leverage the activity data points for all assays (which remain under the control of their respective owner) and as much of the further available side information for compounds (images and transcriptional profiles) as possible. Methods within the scope of this topic should be compatible with the full scale and richness and with the limitations of the above data. For example, given the absence of assay meta-information, no predictive performance gains can be realised by constructing models across data columns with similar or shared annotations.

Predictive performance improvements from federated learning are expected to stem from the multi-task effect across partners. In the rich data sets described above, most assays are poised to show some linear or non-linear correlation with (a combination of) other assays. In a multi-task setting, this allows the model predictivity to be boosted for chemistry that was not documented in the training set for a given assay, but that was documented in some correlated assay(s). In a federated learning setting, such information transfer will occur across partners, through common representation of tasks/assays in federated (as in shared among the data participants) model components. Privacy preservation on the other hand implies that each pharma owner/contributor of assay data builds up (on IT infrastructures under his own control) complementing model components that are specific for his own assays. Federated and privacy-preserving learning combines federated model components (enabling transfer learning across partners) and private model components (to preserve the confidential nature of the modelled assays) to yield better informed, yet overall private models for the respective data owners/contributors. This combination of better learning with preserving the privacy of the underlying data and assets is the core value of the proposed platform.

In terms of predictive performance, the concrete outcome of the evaluation of the platform will be relevant metrics of the predictive performance of the platform as a function of design and setup choices, aggregated by the platform across all the assays from all the partners. Platform-mediated aggregation ensures that contributions of the individual participants to the overall performance are anonymized, in order that here too, privacy is preserved. The aggregated performance metrics will be shared with the consortium partners to guide and improve design choices, and ultimately document the predictive performance of the final versions of platform, a key objective of the proposal, without however disclosing the underlying confidential data and/or any predictive models derived therefrom.

For future exploitation, platform versions must be designed that can also produce the individual predictive models for the assays of respective data contributors, in a form that persists after completion of the run. It should be noted, however, that the generation of such persistent individual predictive models, which are inseparably linked to the private compound and activity data from the EFPIA partners, is not essential for the computation of the aggregated performance metrics during the cross-pharma evaluation rounds. Indeed, these aggregated performance metrics can also be computed using alternative platform versions that do not produce persistent individual predictive models, but this would burden the consortium with the development and audit of an alternative version for each platform iteration. It is crucial to understand that the preservation of privacy and confidentiality of the data to be learned from and the individual models derived therefrom is a key component in the successful implementation of the topic, not only in the current context of discovery and preclinical research but also for any potential future extensions using clinical data. It is also a condition for the involvement of the extensive private datasets of the EFPIA partners.

Technically, privacy preservation is interpreted to exclude any persistent or non-persistent consolidation of assay data or annotations, or the corresponding predictive models, which were described above as the private model components (even encrypted) outside of IT architectures under direct and sole control of their respective owners. It also implies the confidential treatment of all related data and protection from access to them by third parties.

The proposed project aims for federated machine learning which is not the same thing as machine learning on federated data. The difference is as follows: in the former case, the machine learning effort itself is distributed over the parties involved; in the latter case, the machine learning is executed centrally over federated data, which is incompatible with the proposed interpretation of privacy preservation. Upon completion of a modelling exercise, no data (derived or otherwise) should persist outside of those architectures. The pharma IT departments will consolidate their IT security requirements, including those covering compatible cloud services proposed as part of the platform IT architecture, based on current industry standards that aim to protect against illegitimate access to or use of the data or predictive models.

The expected time and cost efficiency gains in a development context (using clinical data) will most likely far outweigh those in the current discovery setting, given the obvious privacy considerations concerning clinical data. It is, therefore, important that the platform is designed with future use in a clinical setting in mind. However, this project focuses on the core objective of developing the federated, privacy-preserving machine learning method in a preclinical setting. Tackling the complexities of clinical data handling in terms of adequately addressing ownership and privacy legislation implications would take place in a future initiative.

To further bolster the confidence in the proposed methods of the pharmaceutical partners (and of potential other future adopters), an intrinsic part of the proposal should focus on analysing the privacy preservation of the proposed methods in the case of legitimate use (targeting questions like “can a model owner reconstruct parts of the chemical or bioactivity data of individual other parties based on model components they can legitimately access”). Public data (prepared and processed by the pharma partners using the same protocols as for their own data) can be leveraged to this end.

In summary, the power of the proposed federated and privacy-preserving machine learning platform resides in the fact that it operates in such a way that it can extract a maximum of learnings without the need to directly access the underlying private data. This makes the methodology generic and widely applicable in a great diversity of settings, with a high potential in settings where learnings are envisaged from highly confidential data, such as patient related data in a clinical setting.

Expected key deliverables

- An early software prototype for federated learning compatible with privacy preservation (not enterprise ready) is delivered by month 2 to allow the algorithm to be documented and to enable an analysis of privacy preservation by the use of legitimate modelling results. This prototype should be based on software already existing at project start.
- A coherent, federated, privacy-preserving machine learning platform that conforms with the following requirements should be delivered by month 12 and updated at least annually.
    - For each iteration, an early software prototype is made available 10-12 months ahead of the enterprise-ready release, to allow algorithm documentation and to enable an analysis of privacy preservation.
    - For each iteration, a report on the privacy preservation performance of the platform using public data, listing algorithmic or parameter options to navigate performance/privacy trade-offs, is prepared. This includes evaluating vulnerabilities to e.g. differential attacks. These reports will enable conceptual sign-off for use on the massive proprietary and confidential pharma datasets;
    - For each iteration, based on the signed-off conclusions of the privacy preservation report, enterprise ready code is delivered, i.e. ready for independent code audit against joint pharma security requirements (that should preclude to reasonable standards illegitimate access to or use of data or models, and that cover compatible IT architecture options including cloud services). A favourable audit report is a prerequisite for exposure of the massive pharma datasets;
    - Ability to be run on a requirements-compatible IT architecture in a standalone and federated learning setting;
    - From the 2nd year onwards, the solutions should enable participants to mutually benefit from the inclusion of high throughput image or transcriptional datasets annotating sets of more than 100k compounds

- Establishment of proof-of-concept of this platform, by deploying and evaluating it in an industrial setting:
    - EFPIA partners to propose common protocols to standardise, format and normalise their private data for optimal interoperability and release software needed to do so;
    - EFPIA partners to consolidate their necessary and sufficient IT security requirements including those covering compatible cloud services proposed by the applicant consortium as part of the platform IT architecture. (Part of) the infrastructure will need to be under the control of the respective data and asset owners;
    - To evaluate the predictive performance of the platform when deployed on industrial scale datasets as a function of design and setup choices, by performance metrics aggregated by the platform across all assays and partners. Relevant performance metrics to include established metrics that can be used with annoted compounds sets, like the AUC of the ROC curve and logarithmic loss metrics or root mean squared prediction error for all assays, aggregated as distributions across all assays and partners. In addition, performance metrics are to be collected, in an aggregated modality, that measure the information gain (i.e. certainty, credibility or precision gain) over the platform of predictions for unannotated compounds.
    - Standalone and cross partner runs yielding these performance metrics (aggregated by the platform across all assays and partners) to be executed on a requirements-compatible ICT infrastructure and comparison of the resulting aggregated metrics. The algorithmic, software and ICT infrastructure choices proposed should cost-efficiently enable a full cross-partner run to complete in maximally four weeks. This may or may not imply provision of hardware acceleration options, and - to ensure availability of such options for all participants - cloud services;
    - At least in one exercise, the aggregated predictive performance of inclusion vs. exclusion of image-derived or transcriptional features in a federated modelling run are compared head-to-head;
    - At least in one exercise, the aggregated predictive performance of the developed methodology is compared head-to-head to that of a credible established non-federated single-task method (minimally support vector machine (SVM), random forest or a comparably performant method);
- Sustainability plans that detail how the applicant consortium intends to make the developed methodologies accessible to the pharmaceutical industry and to other future adopters after the project ends;
- Publication and dissemination of guidelines, advice, detailed processes (workflows and specific technical details), ICT and security standards, and of the predictive performance (at an aggregated level) to promote the uptake of the developed methodologies in the pharma and other) sectors;
- Identification and publication of any barriers to the uptake of the proposed methodology and publication of solutions to reduce those barriers.

Expected impact

The *in silico* predictions from the platform developed within the project will increasingly replace the costly and time-consuming *in vitro* testing, resulting in cost and time savings on compound synthesis and measurement in assays and preclinical studies, and therefore increase the efficiency of pharmaceutical discovery research. Although out of the direct scope of the present topic, the application of similar concepts to clinical data to enable faster recruitment of more targeted patients holds the longer-term promise of reducing costs of development.

The concepts developed within the project will be generic and will apply not only to the pharmaceutical discovery and clinical development setting, but also to other clinical applications, including Real World Evidence analysis. Beyond the health area, they will prove relevant to multiple alternative industrial and other commercial or non-commercial settings where parties are interested in different predictive models that benefit from indirect access to the same volumes of private data. By providing data owners with the confidence that their data and the corresponding predictive models will remain private, this project will facilitate access to much larger data sets and therefore improve performance over that of conventional machine learning approaches

For knowledge and ICT partners, federated learning presents a line of research and product development beyond that of data federation.

Applicants should indicate how their proposal will impact on the competitiveness and industrial leadership of Europe by, for example engaging suitable small and medium sized enterprises (SMEs).

Potential synergies with existing consortia

Applicants should take into consideration, while preparing their short proposal, relevant national, European (both research projects as well as research infrastructure initiatives), and non-European initiatives. Synergies and complementarities should be considered in order to incorporate past achievements, available data and lessons learnt where possible, thus avoiding unnecessary overlap and duplication of efforts and funding.

For example, several IMI projects have already faced the challenge of facilitating research on private data, see http://www.sciencedirect.com/science/article/pii/S1359644615004249 and http://www.mdpi.com/1422-0067/15/11/21136/html

Another IMI project aims at the systematic FAIRification of data (the capture and management of data to make them Findable, Accessible, Interoperable and Reusable). The project consortium is encouraged to seek synergies with projects for the FAIRification of data (e.g. consider applying learnings and technologies from such projects), but should avoid replication of such efforts.

Industry consortium

Key contributions from EFPIA partners:

- Agreed protocols and solutions for processing data with the necessary and sufficient level of standardisation to enable the machine learning exercises. To encourage broader adoption, the partners will opt for open solutions where possible. Insights on data standards and technologies from ongoing EU funded projects (e.g. those in the context of the FAIRification IMI topic) will be considered.
- The anticipated collective industry datasets outlined under Scope, above;
- Data management;
- Formulation of joint security requirements in line with industry standards;
- Set up independent audit of all enterprise-readied code against those requirements;
- Evaluation of the analysis of privacy preservation based on legitimately accessed models;
- Expertise in cheminformatics and machine learning at scale in the context of this topic;
- Upon enablement by the consortium (access to secure software solutions), execute provided solutions on own data (standalone);
- Evaluate the aggregated predictive performance in terms of accuracy and related metrics (for annotated compounds) and information gain and precision (for unannotated compounds);
- Extensive experience in drug discovery and development, including knowledge, of all in vitro and preclinical assays modelled;
- Expertise in image and omics analysis, to facilitate the accommodation of image or transcriptional information in the developed methods;
- Project management coordination across pharma;
- Project management support by a subcontracted project management office;
- Dissemination activities within the sector.

Indicative duration of the action

The indicative duration of the action is 36 months.

Future Project Expansion

Potential applicants must be aware that the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking, may publish at a later stage another call for proposals restricted to those projects already selected under this call in order to enhance and progress their results and achievements by extending their duration and funding. Consortia will be

entitled to open to other beneficiaries as they see fit. If proof-of-concept in terms of privacy and predictive performance is established in the discovery setting, there is the possibility of a restricted call that would adapt the platform developed under the present call for use on clinical datasets, i.e. deliver and evaluate an extended version of the platform that would:

1. map relevant clinical concepts to specific platform components; and

2. meet all additional legal requirements associated with the handling of patient data.(e.g. those related to the protection of patient privacy)..

Applicant consortium

The applicant consortium will be selected on the basis of the submitted short proposals.

The applicant consortium is expected to address all the research objectives and make key contributions to the defined deliverables in synergy with the industry consortium which will join the selected applicant consortium in preparation of the full proposal for stage 2. Therefore, the applicant consortium should be able to demonstrate the full scope of experience and expertise needed to effectively address all the objectives outlined in this topic.The size of the applicant consortia should reflect the expertise needed to achieve the proposed objectives within the indicated budget while ensuring the ‘’manageability” of the consortium as well as efficient and effective team work. Therefore, the number of members of the applicant consortium needs to be thoroughly justified in the proposal and all partners involved should make a significant contribution to the project.

To meet the ambitions of the topic and ensure a first version can be deployed by the end of year one, the applicant consortium should describe the workhorse algorithms they intend to use in their short proposal, convincingly demonstrating their compatibility with the type of data made available for this topic and with the proposed federated and privacy-preserving machine learning concepts, preferably with a (not necessarily secure or enterprise ready yet) software prototype.

Given the runs will involve the handling of private preclinical data sets at an unprecedented scale, the applicant consortium is expected to mobilise across academia, SMEs and other commercial organisations as appropriate, the following

- Demonstrated extensive hands-on expertise in solutions for big data handling at industrial scale;
- Demonstrated extensive hands-on expertise in ICT security and information leakage aspects;
- Demonstrated extensive hands-on expertise with deployment on high performance computing infrastructures;
- Demonstrated extensive hands-on expertise in software engineering;
- Demonstrated extensive hands-on expertise in machine learning technologies, including in the context of federated learning;
- Demonstrated hands-on expertise of deploying computational approaches in the context of drug design, drug discovery and development;
- Demonstrated hands-on expertise in general project management (ability to consistently set and achieve milestones on time and within budget; managing varying interests of multiple stakeholders) and professional communication (expertise in communication tools and systems for project management purposes), in the context of EU-funded projects.

The short proposal should include a description as to how the applicant consortium intends to make the developed methodologies accessible to the pharmaceutical and other industries after the project ends. To this end, it is suggested to allocate responsibility for ensuring sustainability (including software, licensing, infrastructure options, potential broker services) to a specific consortium partner. While a broker role is acceptable, and could for example be filled by an SME, this role must be compatible with the outlined interpretation of federated and privacy-preserving machine learning, for instance the broker function will not have access to assay data, annotation or the corresponding models.

Suggested architecture of the full proposal

The applicant consortium should submit a short proposal which includes their suggestions for creating a full proposal architecture, taking into consideration the Industry consortium contributions and expertise provided below.

The consortium is expected to have a strategy on the translation of the relevant project outputs into regulatory practices, regulatory, clinical and healthcare practice. A plan for interactions with regulatory agencies/health technology assessment bodies with relevant milestones and resources allocated should be proposed to ensure e.g. qualification advice on the proposed methods for novel methodologies for drug development.

In the spirit of the partnership, and to reflect how IMI2 JU call topics are built on identified scientific priorities agreed together with EFPIA beneficiaries/large industrial beneficiaries, these beneficiaries intend to significantly contribute to the programme and project leadership as well as project financial management.

The final architecture of the full proposal will be defined by the participants in compliance with the IMI2 JU rules and with a view to the achievement of the project objectives. The allocation of a leading role within the consortium will be discussed in the course of the drafting of the full proposal to be submitted at stage 2. To facilitate the formation of the final consortium, until the roles are formally appointed through the consortium agreement, the proposed project leader from among EFPIA beneficiaries/large industrial beneficiaries shall facilitate an efficient negotiation of project content and required agreements. All beneficiaries are encouraged to discuss the project architecture and governance and the weighting of responsibilities and priorities therein.

A plan for aspects related to sustainability, facilitating continuation beyond the duration of the project should also be proposed.

The architecture outlined below for the full proposal is a suggestion. Different innovative project designs are welcome, if properly justified.

Work package 1 – Preprocessing of data up to a level of necessary and sufficient standardisation

- Select methodology for standardised preprocessing data and implement in scripts, including feature extraction, dimensionality reduction, weighted data integration;
- Enable participants to deploy scripts in standardised ways compatible with the architectures proposed for the exercise;
- Execute preprocessing of data and making them available (including public data for work package 3).

Industry consortium contribution:

Methodology selection, implementation and execution.

Expected applicant consortium contribution:

Enable architecture-compatible deployment, scientific advice.

Work package 2 – Industrial IT technical scoping and deployment

The goals of this work package will be as follows:

- Joint pharma user requirements;
- Independent software audit (in kind pharma contribution) of the resulting software (from work package 5);
- Enable/execute runs on ICT infrastructure under pharma control (these may be cloud services).

Industry consortium contribution:

Formulation of user requirements, set-up of audit, enable runs.

Expected applicant consortium contribution:

Liaison between pharma driven work package 2 and consortium driven WP5 (software implementation), to ensure solutions match requirements and can be run on pharma controlled infrastructures.

Work package 3 Federated Machine Learning Algorithms

The goal of this work package will be as follows:

- Development and scientific and software prototyping of the algorithm;
- Initial predictive performance estimation (on public data);
- Machine learning security analysis of algorithms (on public data), to enable security evaluation.

Industry consortium contribution:

Experts in machine learning applied to the domain of the topic.

Expected applicant consortium contribution:

Expertise to carry out the activities listed above.

Work package 4 – Evaluation of privacy and performance balance and of predictive performance of the versions up to implementation in discovery projects

The goals of this work package will be as follows:

- Evaluation of balance between performance and privacy preservation (on prototypes);
- Evaluation in terms of the aggregated predictive performance metrics (enterprise-ready product)

Industry consortium contribution:

Expertise to carry out the activities listed above.

Expected applicant consortium contribution:

Scientific support for activities listed above.

Work package 5 – Software Implementation

The goals of this work package will be as follows:

- balance in WP4 (scientific) and WP2 (data privacy), to be readied to the point that it can be securely deployed on the massive pharma datasets;
- This includes aspects of software engineering, ICT security, knowledge of ICT infrastructure to run on, with respect to software implications (high performance computation enablement, hardware acceleration, …).

Industry consortium contribution:

Industrial experts in ICT, security, machine learners and modelling.

Expected applicant consortium contribution:

Expertise to carry out the activities listed above.

Work package 6 – Secure Standalone and Federated Infrastructure

The goal of this work package will be as follows:

- Provision of infrastructure that will operate under control of the respective EFPIA data and asset owners during standalone and federated runs (may be cloud services)
- Provision of central ICT infrastructure that can connect to the infrastructures under control of the respective EFPIA data and asset owners involved, ensuring security and performance requirements;
- Operation Support.

Industry consortium contribution:

Industrial experts in ICT.

Expected applicant consortium contribution:

Selecting, setting up and providing the secure infrastructure for standalone and federated modelling runs to be procured under the action.

Work package 7 - Operations and Deployment

The goals of this work package will be as follows:

- Establish a detailed software and operating model with pharma organisations;
- Monitoring execution of runs upon initiation by pharma.

Industry consortium contribution:

Industrial experts in ICT and modelling.

Expected applicant consortium contribution:

- Main drivers, may include partners involved in sustainability plans

Work package 8 - Overall project governance, project management, dissemination and sustainability

The goals of this work package will be as follows:

- Grant administration;
- Strategic, operational, IP and financial management;
- Communication (within the consortium and with relevant external collaborators);
- Dissemination of scientific results and research data to the scientific community and within the pharma sector;
- Detailed sustainability plan to make results accessible beyond the duration of the action.

Industry consortium contribution:

Programme leadership with respect to application and valorisation aspects, project and financial management, contribution to communication and dissemination.

Expected applicant consortium contribution:

Scientific and technical programme coordination, reporting to the IMI2 JU (supported by the industry-provided project management expertise and support).