Standardisation in life-science research - Making the case for harmonization to improve communication and sharing of data amongst researchers.

Susanne Hollmann[*1, 2], Babette Regierer[1], Domenica D'Elia[3], Kristina Gruden[4], Špela Baebler[4], Marcus Frohme[5], Juliane Pfeil[5], Ugur Sezerman[6], Chris T Evelo[7], Friederike Ehrhart[7], Berthold Huppertz[8], Erik Bongcam-Rudloff[9], Christophe Trefois[10], Aleksandra Gruca[11], Deborah A. Duca[12], Gianni Colotti[13], Roxana Merino-Martinez[14], Christos Ouzounis[15], Feng He[16], Oliver Hunewald[16], Andreas Kremer[17].

[1] SB Science Management UG (haftungsbeschränkt), 12163 Berlin, Germany
[2] Research Centre for Plant Genomics and Systems Biology, Potsdam University, Potsdam 14476, Germany
[3] CNR - Institute for Biomedical Technologies for Biomedical technologies, Bari, Italy
[4] Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia
[5] Technical University of Applied Sciences, 15745 Wildau, Germany
[6] Acibadem University, 34750 Istanbul, Turkey
[7] Dept. Bioinformatics, Maastricht University, The Netherlands
[8] Dept. Cell Biology, Histology & Embryology, Gottfried Schatz Research Center, Medical University of Graz, Graz, Austria
[9] Dept. Animal Breeding and Genetics, Bioinformatics section. Swedish University of Agricultural Sciences, Uppsala, Sweden
[10] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg
[11] Institute of Informatics, Silesian University of Technology, Poland
[12] Research Support Services, University of Malta, Malta
[13] Italian National Research Council, Institute of Molecular Biology and Pathology, Rome, Italy
[14] Karolinska Institutet, Stockholm, Sweden
[15] Centre for Research and Technology Hellas - CERTH, Thessaloniki, Greece
[16] Department of Infection and Immunity, Luxembourg Institute of Health, L-4354 Esch-sur-Alzette, Luxembourg
[17] ITTM S.A., 4354 Esch-sur-Alzette, Luxembourg

Modern, high-throughput methods for the analysis of genetic information, gene and metabolic products and their interactions offer new opportunities to gain comprehensive information on life processes. The data and knowledge generated open diverse application possibilities with enormous innovation potential. To unlock that potential skills in generating but also properly annotating the data for further data integration and analysis are needed. The data need to be made computer readable and interoperable to allow integration with existing knowledge leading to actionable biological insights. To achieve this, we need common standards and standard operating procedures as well as workflows that allow the combination of data across standards. Currently, there is a lack of experts who understand the principles and possess knowledge of the principles and relevant tools. This is a major barrier hindering the implementation of FAIR (findable, accessible, interoperable and reusable) data principles and the actual

reusability of data. This is mainly due to insufficient and unequal education of the scientists and other stakeholders involved in producing and handling big data in life science that is inherently varied and complex in nature, and large in volume. Due to the interdisciplinary nature of life science research, education within this field faces numerous hurdles including institutional barriers, lack of local availability of all required expertise, as well as lack of appropriate teaching material and appropriate adaptation of curricula.

Keywords: FAIR data, Standardisation, Interoperability, Standard Operating Procedures (SOPs), Quality Management (QM), Quality Control (QC), Education.

Introduction

Life sciences are undergoing major changes in research practice, with cross-cutting, global efforts to maximise the benefit of research outputs for all members of the scientific community. Key to these changes is the recognition of the importance of data, appropriately annotated. Long term accessibility of research data is relevant not only to reproduce results but also as a starting point for reuse in other studies. Publication of data therefore is more and more considered a primary output. Consequently, today, researchers must articulate data-sharing, -preservation and -publication strategies at the point of planning their projects, and robust data management plans have become mandatory components of funding proposals such as within the European Framework Programme H2020 [1].

The challenges of providing reliable research data accompanied the scientific community since the beginning of research activities. In 2005 the RDM (Research Data Management) initiative introduced a new set of principles for data management services. Following their principles, data should be FAIR – Findable, Accessible, Interoperable and Reusable [2]. The main objective of the FAIR Data Principles is the optimal preparation of research data for man and machine. In these days following the FAIR Data Principles in research is almost a condition to receive funding even if it is not always accompanied with a positive response from all involved stakeholders.

However, research can only evaluate the enormous amounts of data to a limited extent. A closer cooperation between the life sciences and other subjects such as mathematics, computer science and engineering are needed to cope with this "Big Data" and, of course, their direct engagement in bioinformatics. To facilitate knowledge gain by assisting us in the discovery chain including access of data, their integration and analysis, and their processing with associated algorithms and workflows, the lack of knowledge of the relevant tools and skills so far needs to be filled by intense collaboration with other fields and disciplines.

The value of data stewardship is universally recognised, currently more in principle than in practice: some £3 billion of public money is annually invested in research in the UK alone, yet the data resulting from this considerable investment are seldom as accessible as they could or should be (http://www.dcc.ac.uk/digital-curation). The German Research Foundation (DFG) estimates that 80-90% of all research data are never shared with other researchers, but remain unpublished, often hidden away in a drawer or on an individual's hard-drive [3]. Thus, research data are lost due to unacceptable storage conditions, outdated or unusable formats, missing backup, incomplete or unpublished work and finally, even if published, by inefficient documentation and, most importantly, by a certain lack of incentives. This loss of data caused extremely high costs because it required the repeated use of resources, time and manpower and was hindering a stable advancement of knowledge and innovation. This issue crosses many disciplines and affects basic as well as preclinical research [4]. The implementation of open access publications and FAIR data principles within the European Framework programs has been a logical consequence and the first step to overcome these hurdles. The preservation and sharing of digital material for reuse by others maximises the impact of research and inspires confidence in the research councils and funding bodies that invest in the work. Furthermore, it is now widely recognized that making research results more accessible to all societal actors contributes to better and more efficient science, and to innovation in the public and private sectors [5].

The cost of having to repeat poorly conceived experiments is high; the cost of publishing false or unusable data is potentially more far-reaching and damaging (to the individual scientist, the research community and, indirectly, to society at large), either because errors are, in this way, propagated 'in the wild', or because poor formatting means 'good' data cannot be utilised. Therefore, high-quality information is imperative before experimental design commences; and, for this, education and access to tools and training materials to build awareness of, and proficiency in, Open Science and FAIR principles are essential. For example, researchers need to understand the importance of using standard identifiers to unambiguously refer to biological and chemical entities so that the contents of their data-sets, and of the papers that refer to them, can be described and/or linked to core biological/chemical databases in computer- accessible ways. As scientific data varies greatly we have to consider that there will be not one standard to fit all but many of them. Additionally, these will be adapted over time as methodologies and requirements change. Providing tools that help scientists to do this and providing training on how to achieve rigorous and consistent identification is a fundamental step towards making research data FAIR. On the other side of the coin, publishers have an obligation to ensure the integrity of the research they publish; this means that they too have a responsibility to develop guidelines for data-management, linked to FAIR principles, and to help researchers to understand and meet their data-stewardship standards.

Driven by the large-scale approaches of systems biology, the need for implementation of FAIR principles and availability of data has been recognised already years ago with the result that repositories, platforms and tools already exist for the community and respective training offers have been developed to use these excellent resources. These initiatives represent an excellent basis to go beyond and achieve a dissemination on a broader scale within the life science community and the other relevant stakeholder groups.

However, there are still some barriers to overcome:

Metadata vs. raw data: The FAIR data principles, mainly address the metadata levels in research, whereas the quality of the source datasets themselves often remains unaddressed. Even if the datasets are published following the FAIR data principles, the quality of the actual data might be unsatisfactory. As a result, downstream calculations, analysis and proceedings based on such data might lead to questionable reproducibility.

High quality for samples and data: An essential prerequisite of modern life-science R&D is a high quality of the research data. By enabling the reuse of research assets, research becomes considerably more efficient and economical. This can only be achieved reliably and efficiently if these are generated according to standards, using Standard Operating Procedures (SOPs), are managed according to a Data Management Plan and are hosted on sustainable infrastructures. The data quality is directly linked to the quality of the biological samples used. Hence, high-quality data can only be obtained by the respective use of high-quality samples [6].

Low acceptability caused by lack of education: According to Springer Nature Survey of more than 7,700 researchers worldwide, 76% of responders highly rate the importance of their data being discoverable with the average rating 7.6 of 10 [7]. Interestingly, the biological sciences had the highest proportion of respondents who share data relating to publications (75%), followed by the Earth sciences (68%), medical sciences (61%), and physical sciences (59%). However, in the same survey, almost half of the researchers (46%) admitted that organising data was a challenge, followed by confusion around copyright (37%), not knowing where to share the data (33%), lack of time to deposit data (26%) and finally costs of sharing data (19%). The issues related to lack of knowledge, which repository to use and uncertainty about copyright and licensing were particularly seen as problems for early career researchers. This indicates that improving education and support on good data management could increase data sharing, and thus reusability in the scientific community.

Good laboratory practices imply mandatory, well-defined and precisely described techniques, methods and protocols towards optimal and reproducible conditions. This leads to interlaboratory comparability, reproducibility of experiments and reusability of data and obtained scientific results.

In addition, as the study above shows, proper data management and established analytics workflows must be put in place. Similarly, to laboratory scientists, computational scientists should consider establishing data analysis related SOPs and workflows. Furthermore, the use of electronic lab notebooks is encouraged as it allows proper documentation, traceability and sharing of information between researchers [8]. Research infrastructure can be instrumental in helping scientists as well as the scientific community by providing structured environments. In such environments, standards and SOPs should be applied and respected contributing by providing relevant platforms for data acquisition, handling, storage and/or analysis. These platforms can be a combination of instruments, storage capabilities, high performance computing, training opportunities, databases and others. Some well-known infrastructures in life sciences are ELIXIR, BBMRI-ERIC and COBEL, to name a few (ESFRI, 2017).

Thus, standards and research infrastructures represent important drivers in the life-sciences and technology transfer because they aim to guarantee that data become accessible, shareable and comparable throughout their lifetime.

The need for proper standards and tools pushed several grassroot initiatives to develop standards and frameworks for their implementation in the past. For instance, the obligation to publish genetic data together with the research paper by the journal Nucleic Acid Research (see their Instructions to Authors). Unfortunately, these efforts remain fragmented and largely disconnected [9]. These initiatives put their focus mainly on the interaction and exchange with experienced researchers within their field, but do not necessarily take into account to involve the non-expert researchers in the training and education aspects of their actions.

In addition, despite the range of local, national and transnational training and education activities, many students and researchers in Europe do not have an easy access to good quality systems biology education and need better access to high-quality systems biology education and training.

In order to achieve a consistent and reliable system for producing high-quality scientific data, we propose several approaches:

1) introducing certified methods and protocols in a top-down process. However, the procedures imposed in such a way are certain to be received by a community as another unnecessary administrative element. There is also a debate on which institutions are qualified to decide which methods and protocols should be implemented and to what extent. It may be very difficult or even impossible to define uniform quality seals that could be used and recognised across different scientific communities.
2) encouraging a EU-wide adoption of training of young scientists.
3) provide open access protocols, methods and tools serving as a basis for proper data management.
4) encourage publishers to request authors to make the source data available for any publication they consider.

Common training and education across Europe

Modern high-throughput methods for the analysis of genetic information, proteins and metabolic products offer new opportunities to gain comprehensive data on life processes. The respective results open diverse application possibilities with enormous innovation potential. To facilitate knowledge gain by assisting in the discovery chain including access of data, their integration and analysis, and their processing with associated algorithms and workflows, the scientists experience lack of knowledge on data management as well as accessibility of the relevant tools. This is mainly due to the lack of common standards and standard operation procedures and to an insufficient and unequal education of the scientists and other stakeholders involved in producing and dealing with these big data. As life science research is inherently multidisciplinary, education within this field meets numerous hurdles including departmental barriers, availability of all required expertise locally, appropriate teaching material and example curricula.

As university education at the bachelor's level is traditionally built upon disciplinary degrees, we believe that the most effective way to implement education in Data Management, Standardisation and Quality Management should start at the bachelor's level and continue on the Master's level, as it offers a more flexible framework. Within the Cost Action CHARME (Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research – CA 15110) participating experts and active performers of Standards suggest:

(i) a definition of the minimal skills that students should acquire within their Bachelor studies and Master's programme,
(ii) a possible basic educational curriculum with flexibility to adjust to different application areas and local research strengths,
(iii) involvement of teachers, principal investigators and lecturers to increase awareness within this group,
(iv) mechanisms for collaborative preparation and sharing high quality teaching materials and methods among education professionals.

The lack of knowledge is mainly due to the lack of common standards and standard operation procedures in combination with an insufficient and unequal education of the scientists and stakeholders involved in producing and dealing with Big Data. Their unequal education must be improved by proposed approaches. However, the implementation of standards represents the most important driver in the life-sciences and technology transfer because it guarantees that data become accessible, shareable and comparable along the value chain and thus facilitates the daily work processes and in the end accelerate innovation transfer.

Nevertheless, there are many hurdles to overcome in order to reach a level of productive innovation: the challenge of enabling optimal use of methods and its resulting research data is complex and involves multiple stakeholders including:

- Individual researchers
- Funding agencies
- Data scientist communities
- Publishers
- Institutional managements
- Clinicians
- Service providers (e.g. for data generation/analysis, processing and stewardship)
- Industry (e.g. software and tool-builders, pharmaceutical and healthcare industry, biotechnology, chemical industry…)

Each of the groups above has different interests and different needs. However, it is important that all the key stakeholders are aware of each other, working in complementary ways.

Furthermore, looking to Europe and beyond, the available infrastructures and education systems vary enormously. Here, relevant national and international research and innovation activities need to be connected to combine efforts and to avoid duplication and fragmentation. Examples for activities and initiatives relevant on the European level are:

- ELIXIR-TeSS: Training eSupport System, a Portal provided by ELIXIR for training material on bioinformatics (https://tess.elixir-uk.org/))
- FAIRDOM: a platform supporting collecting, managing, storing, and publishing research data, models, and operating procedures (https://fair-dom.org/).
- SBEDU: Platform providing training material for systems biology (http://www.sbedu.eu/)
- ENFIN: integration of tools for systems biology (http://www.enfin.org/), early efforts
- GOBLET: Global Organisation for Bioinformatics Learning, Education and Training (http://www.mygoblet.org/)
- MANTRA: Research Data Management Training (https://mantra.edina.ac.uk/)
- Data carpentry: develops and teaches workshops on the fundamental data skills needed to conduct research with the mission to provide researchers with high-quality, domain-specific training covering the full lifecycle of data-driven research (http://www.datacarpentry.org/).
- Software carpentry: (https://software-carpentry.org/) provides training materials, develops and teaches workshops on basic software skills for researchers
- CHARME: COST Action on Standardisation in the Life Sciences (https://www.cost-charme.eu/)
- IMI eTRIKS: Is the result of a collaboration between 17 different partners. eTRIKS provides advice, open source platforms and training to translational research projects (https://www.etriks.org/)
- IMI TRAIN: Is a partnership between EMTRAIN, Eu2P, PharmaTrain and SafeSciMet and provides courses, training programmes and tools (http://www.imi-train.eu)
- EDISON: is an EU funded project to accelerate the creation of the Data Science profession (http://edison-project.eu)
- FAIRsharing (https://fairsharing.org/), formerly known as BioSharing, an online repository for databases, standards and policies developed *i.a.* by Elixir, EMBLand several journals (Sansone et al. 2018).
- EMBnet: a world-wide organisation that brings bioinformatics professionals together to serve, support and sustain the growing field of Bioinformatics in the Biological and Biomedical research domains (https://www.embnet.org/wp/)

European Infrastructures (ESFRI) [10] such as BBMRI-ERIC (Biobanking and BioMolecular resources Research Infrastructure), ELIXIR (European life science infrastructure for biological information) and ISBE (Infrastructure for Systems Biology Europe) represent relevant information hubs as the infrastructures integrate research institutions (academia and industry) across all European countries. However, even those initiatives sometimes lack in terms of SOPs and miss a sufficient and equal education of the scientists and stakeholders involved.

In this context, fundamental challenges exist that hinder innovation transfer:

at the educational level (individual researchers)
at the institutional level
at the administrative (legal) level and even at the industrial level

The challenges at the educational level:

For early career researchers, the variety of tools and standards available can be overwhelming. The choice and decision of the right tool for inexperienced researchers can be difficult and in the worst case ultimately compromise the quality of research data and generation of proper research results.
To this end, detailed guidelines on how to conduct a research study including guidelines on processing and recording of data will be beneficial for the scientific community. Generally speaking, the research process begins with experimental design (including selection of appropriate tools and resources, formats, analysis methods, choice of data warehouse, etc.) and ends with publication. During this process, each step and each decision forms the basis for the next step. At any stage, poor or wrong decisions can create problems, potentially resulting in entire experiments having to be repeated and/or incorrect or poorly-formatted data being published.
Training scientists to generate, format and curate their own data throughout their education and career will allow them to be ready for both publishing and openly sharing their data in repositories. This would be a significant advance over today's practices. Because of the diversity of tools and resources available across the life sciences, drafting a holistic and efficient structure for training is challenging, albeit an absolute must. Any training strategy must be independent and impartial and should enable researchers to identify the most appropriate tools, platforms and courses for their needs.
Only with a consistent and uniform approach towards education in data stewardship at large, the quality of data, their reproducibility and interoperability can be guaranteed, and the data become usable for all interested parties independent of the origin of their production. At the same time, uniform training in data generation and data management across every involved disciplines needs to be established and implemented.

The challenges at the institutional level:

University rankings are above all an important political instrument. Academic globalization is happening very fast. Universities recruit their employees around the world; scientists commute back and forth between different countries, and investors worldwide are looking for promising research projects. Governments want to have an overview of which research is leading and the establishment of prestigious universities means innovation and economic progress for them. Sometimes, millions of investments are made mostly based on university rankings. A major component for calculating prestige within institutions is the number of publications and their impact factor. The more publications in high ranked journals, the higher the institution will be ranked in principle. One of the most prestigious university rankings in the world, the "Times Higher Education World University Ranking" (https://www.timeshighereducation.com/world-university-rankings) periodically presents the ranking list of the world's leading universities. Criterias are teaching, research and the transfer of knowledge between university and companies. A third of the evaluation score is based on paper citations.

However, publications and actual results exploitation are sometimes incompatible with each other, which means that the evaluation criteria should be jointly rethought and redefined. In some disciplines, such as bioinformatics, the highest ranked journal will not have the same impact factor as the highest ranked journal in medicine. This means that assigning a too high importance to journal ranks might unfairly attribute less importance to research results that may have a huge impact on the scientific community. For this reason, we propose that in addition to the number of publications and citations, the quality of produced data and the impact of the research results on the scientific community should be integrated into such institutional evaluation criteria.

Similar to the implementation of FAIR principles into the research scholarly, seals of data quality and handling should be established. This means the establishment and implementation of a minimal Quality Management System to show and guarantee the reproducibility and the reliability of the data, and approval certificates presenting the high quality of "how the data has been generated". These active measures need to be supported by the hosting institutions, the funding agencies and the governments. Without institutional support, quality efforts will remain a hurdle too high for many researchers to pass.

"*Institutions and committees should give more credit for teaching and mentoring: relying solely on publications in top-tier journals as the benchmark for promotion or grant funding can be misleading and does not recognize the valuable contributions of great mentors, educators and administrators"* [4].

The challenge at the legislative level:

The research process of the future will be global, networked and open. It is estimated that in Europe currently 1.7 million researchers and 70 million professionals contribute to science and innovation (reference?). Many more actors will take part in different ways and the traditional methods of organising and rewarding research will also see many changes. Citizen science is a currently emerging trend and needs to be included in the Open Science efforts. The vision is that all these experts will share and contribute to a virtual environment with open and seamless services for the storage, management, analysis and re-use of data that is linked to their research activities, across borders and scientific disciplines. Nevertheless, data exchange across institutions and across countries is hindered by different national regulatory systems, especially with respect to patient data. For policy-makers, thus, the task is also about how to develop and align new policies and practices to address the problems discussed above and enhance the impact of solutions.

Funding institutions - not only European but also national governmental, industrial and private - should request from the individual researcher as well as from the grant-receiving institution comprehensive QC/QM systems. In a timeline a gradual system to introduce such requirements may start with financial support for the setup of QC/QM systems or benefits for established ones but finally leading to strict requests in funding guidelines to have such mechanisms established.

Within the Cost Action CHARME (*Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research)* [11] scientists and stakeholders from industry, standardisation initiatives and bodies from more than 30 countries work together to harmonize existing formats and develop new solutions. CHARME specially focus on the development of uniform cross border training. During several collaborative meetings and workshops within CHARME and beyond we discussed and developed the following recommendation to which we invite you to further discuss with us and to contribute to the challenges discussed above!

Our recommendations:

- Journal editors must play an active part in initiating a cultural change. Greater dialogue should be encouraged between scientists, funders, industry and politicians;
- Institutions and committees should give more credit for reliability of data, teaching and mentoring: relying solely on publications in top-tier journals as the benchmark for promotion or grant funding must be reassessed. A set of new evaluation criteria for internal and external ranking of the scientists needs to be established which is recognised by all involved stakeholders;
- Funding organizations must recognize and embrace the need for new tools and infrastructures and assist in their development and in providing greater community access to those tools such as large SOP collections with easy investigator access. In parallel, instruments should be implemented that establish new practices that reinforce the common vision;
- Institutions, funders and public bodies must recognise that QM cannot be done by the single researcher – Infrastructures and support for the researchers but also regulations are needed. This can be done by introducing a minimal quality management system and establishment of a set of *Seals of Quality* for wet lab research (*e.g.*, ranking from A to D); The quality labels could be implemented by first inviting institutions to participate on a voluntary basis, but then a mandatory achievement of a specific label after certain number of years should be reached. Restrictions to funding for institutions with lower label would additionally encourage institutions to set this approach. Quality of data and quantity of data should be ranked by a factor such as indication of the quality value (QV) of next-generation-sequencing data. If the Factor is high the quality of data is high too;
- Uniform training and education cross-borders. This can be done by sharing core teaching materials, textbook, ideas and examination criteria in Standards. In addition, regular training of the trainers should be implemented to keep the seal. There is a strong need to establish mechanisms for collaboration and excellence spreading among education performers in Europe and globally. Governments together with universities must play an active part in initiating a dialogue to establish minimal agreements on common programs in education, tools and SOPs.

Acknowledgement
This publication is based upon work from COST Action CHARME (CA15110) supported by COST (European Cooperation in Science and Technology).
www.cost.eu

References

1. H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, last accessed 2018-05-08.
2. Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data* , *3*, 160018.(2016) Winkler-Nees, http://www.ariadne.ac.uk/issue64/datacite-2010-rpt
3. Freedman, L. P., Cockburn, I. M., & Simcoe, T. S. The economics of reproducibility in preclinical research. *PLoS Biology*, *13*(6), 1–9. (2015). https://doi.org/10.1371/journal.pbio.1002165
4. Begley, C. Glenn., Ellis, Lee M. Drug development: Raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012) doi:10.1038/483531a
5. Benefits of sharing. Nature. 530 (7589): 129–129. **2016-02-11**. doi:10.1038/530129a.
6. Doucet M, Becker KF, Björkman J, Bonnet J, Clément B, Daidone MG, Duyckaerts C, Erb G, Haslacher H, Hofman P, Huppertz B, Junot C, Lundeberg J, Metspalu A, Lavitrano M, Litton JE, Moore HM, Morente M, Naimi BY, Oelmueller U, Ollier B, Parodi B, Ruan L, Stanta G, Turano P, Vaught J, Watson P, Wichmann HE, Yuille M, Zaomi M, Zatloukal K, Dagher G. Quality Matters: Annual Conference of the National Infrastructures for Biobanking. Biopreserv Biobank 2017;15(3):270-276. (2016) doi: 10.1089/bio.2016.0053.
7. Stuart D. et al.: Whitepaper: Practical challenges for researchers in data sharing, (2018). DOI: https://doi.org/10.6084/m9.figshare.5975011.v1
8. Kanza, S.; Willoughby, C.; Gibbins, N.; Whitby, R.; Frey, J. G.; Erjavec, J.; Zupančič, K.; Hren, M.; Kovač, K. Electronic Lab Notebooks: Can They Replace Paper? *J Cheminform* , *9* (1), 1–15. (2017)
9. Sansone, S.; Mcquilton, P.; Rocca-serra, P.; Gonzalez-beltran, A.; Izzo, M.; Lister, A.; Thurston, M. FAIRsharing: Working with and for the Community to Describe and Link Data Standards , Repositories and Policies. *bioRxiv* , No. 1, 0–9. (2018)
10. ESFRI. Horizon 2020 and the Research Infrastructures Landscape, 1–14, https://ec.europa.eu/research/infrastructures/pdf/ri_landscape_2017.pdf#view=fit&pagemode=none, last accessed 2018-05-09. (2017).
11. CHARME Homepage http://cost-charme.eu/