Journal of Computational Systems and ICTs

Practices for the development of applications (Apps) that contribute to the generation of data banks for use in Health

Prácticas para el desarrollo de aplicaciones (Apps) que abonen a la generación de bancos de datos para uso en Salud

ORTEGA-LAUREL, Carlos†*, PÉREZ-MARTÍNEZ, Francisco, SIMENTAL-FRANCO, Víctor Amaury and LÓPEZ-MALDONADO, Guillermo

Universidad Autónoma Metropolitana, Department of Information Systems and Communications, Unidad Lerma, Av. de las Garzas No. 10, Col. El Panteón, Lerma de Villada, 52005 Estado de México.

ID 1st Author: *Carlos, Ortega-Laurel* / **ORC ID:** 0000-0001-6072-8480, **CVU CONACYT ID:** 104250

ID 1st Co-author: *Francisco, Pérez-Martínez* / **ORC ID:** 0000-0001-5770-7849, **CVU CONACYT ID:** 161020

ID 2nd Co-author: *Víctor Amaury, Simental-Franco* / **ORC ID:** 0000-0001-8720-9787

ID 3rd Co-author: *Guillermo, López-Maldonado*

Abstract

Nowadays it is perceived necessary that, for the deep understanding, prevention and solution of health problems, the generation and collection of data, regarding their health, of the population in general, and those of the study in specific. Such that such data eventually facilitate decision-making, based on quantitative and qualitative information, to local, regional and / or global health organizations. At the same time, information and communication technologies, and computing, through technologies such as smartphones and applications, present opportunities to facilitate the generation and collection of data, of a diverse nature, except for those concerning health, to be useful for addressing and solving problems related to health, such as social, economic, cultural and health issues. Under these premises, we propose, and consider, the good practices for the development of applications (Apps), based on the experience documented in the literature, and our own, in the development of applications, such that these experiences enable the generation and data collection related to health, such that, by following them in app development projects, it becomes possible to have data in a standardized way, so that in the future they are analyzed and transformed into useful, accessible and understandable information for the benefit of the health of human beings. This is because local, state, federal and global organizations, such as the World Health Organization, seek to have data on diseases, such as, to name a few examples: non-communicable (chronic) Diseases (CNCDs) and Mental Health (NMH) or others.

Applications, Smartphones, Databases, Health, Data, Programs, Information

Resumen

Hoy día se percibe necesario que, para la comprensión profunda, prevención y solución de problemáticas de salud, se demanda la generación y acopio de datos, en cuanto a su salud, de la población en lo general, y los de la en estudio en lo específico. Tal que dichos datos eventualmente faciliten la toma de decisiones, con base en información cuantitativa y cualitativa, a los organismos de salud locales, regionales y/o mundiales. A la par, las tecnologías de la información y comunicaciones, y la informática, a través de tecnologías como los teléfonos inteligentes y las aplicaciones, presentan oportunidades para facilitar la generación y acopio de datos, de naturaleza diversa, no siendo la excepción los concernientes a la salud, con la finalidad de ser útiles para el abordaje y la solución de problemáticas vinculadas a la salud, como lo son las sociales, económicas, culturales y las propias de salud. Bajo estas premisas, proponemos, y se plantea, algunas buenas prácticas para el desarrollo de aplicaciones (Apps), con base en la experiencia documentada en la literatura, y la propia, en el desarrollo de aplicaciones, tal que dichas experiencias habiliten la generación y acopio de datos relacionados con la salud, de manera que, al seguir u orientarse con las mismas en proyectos de desarrollo de apps, se alcance a obtener y/o conservar datos de manera normalizada, para que en el futuro se analicen y transformen en información útil, accesible y comprensible en beneficio de la salud de la población y los seres humanos en general. Esto en atención a que organismos locales, estatales, federales y mundiales, debieran, tal como sí lo hace la Organización Mundial de la Salud, buscar tener datos sobre enfermedades, como pueden ser, por mencionar un par de ejemplos: las Enfermedades (crónicas) No Transmisibles (ECNT) y las de Salud Mental (NMH por sus siglas en inglés) u otras.

Aplicaciones, Teléfonos inteligentes, Bancos de datos, salud, Datos, Programas, Información

*Correspondence to Author (E-Mail: c.ortega@correo.ler.uam.mx)
† Researcher contributing as first author.

Introduction

As part of a study of smartphone penetration worldwide, for the period 2016-2020, it was found that, as a percentage of the world's population, the global smartphone penetration rate was predicted to exceed 44 percent for the first time by the end of 2020. With 3.2 billion smartphone users worldwide and a global population of approximately 7.7 billion, global smartphone penetration by the end of 2020 is thought to reach 44.9 percent, which implies that today about 44.9 percent of cell phones in use around the globe are precisely smartphones or smartphones, which opens up and enables further expansion in the use of applications (Apps) developed for specific purposes. This is derived from the fact that, in addition to enabling the use of telephony, smartphones integrate both a user-friendly graphical interface and a variety of sensors available to perform different measurements or metrics for specific purposes (Statista, 2020), (O'Dea, 2020). In this sense, a number of potential applications are opened, to make use of the diversity of available sensors, among which is the development of applications for health purposes, i.e., to propose them for medical and clinical use. For example, as an example of the present research, there is the study of Non-Communicable (chronic) diseases (NCDs) and Mental Health (NMH) (World Health Organization, 2020).

In fact, it is relevant to mention that the World Health Organization maintains an "organic group" of Noncommunicable (chronic) diseases (NCDs) and Mental Health (MHD), to cover and follow up international actions on NCDs and mental health disorders, in addition to others such as malnutrition, violence and injuries, and disabilities. Thus, for the WHO to be able to cover and monitor international actions, it requires hard, raw and processed data that are information for the purposes it has set (World Health Organization, 2019).

Nowadays, when the development of an application or system is considered, the specifications are universally thought of in terms of what the user who is requesting the development demands, covering the user's expectations in terms of data and functionalities, building it, in the best of cases, based on standardized development practices and, if not, in terms of the experience of the person in charge of the task, at the free will of the developer.

While this may be the case for general purpose systems or specific purpose proprietary systems that generate their value through such systems, because they seek to meet an information need of equal intention, when it comes to applications or information systems such as the one in question, that is, to meet the demand for the generation and collection of data, This has to change, because the data generated or collected in health-related systems should be put to use for the prospection of the health conditions of the populations in which they are obtained. For this reason, it is necessary to think about the use and management of data, the governance-management of data, which is necessary so that, when viewed in an integral manner, as an ecosystem in which the asset is data, the governance-management model structures the proper functioning of the health data ecosystem.

However, given the connectivity and sensitivity of the data to be processed (health data), it is possible to foresee the potential malicious intention of compromising them, which would result in not sharing them, so it is necessary to generate management-structure mechanisms to articulate the task without violating or compromising privacy. Therefore, a "governance-management model" is proposed for this purpose, which will become part of the "good practices" that will also be outlined.

Under these premises, this paper proposes to conceptualize a model that enables the availability of data, but preserves privacy, in order to capture the information of populations to be studied, in general or specific populations, through applications (Apps) and "good practices" to put the model into use. As an example to show the usefulness of the proposed model, the assumption of use for the collection of NCD and NHM data is presented, from the diagnostic approach with psychiatric instruments (batteries) applied via App. This is because it is clear that questionnaires can be answered from any smartphone, either to obtain diagnoses and/or to generate data that can be collected, with the intention of having present opinions, or to form databases with statistics (historical databases, with specific regularities), which enable the necessary input to conduct in-depth research of NCDs and NHM, which are mentioned as an example.

This would lead to the generation of massive databases that in turn would allow searching for possible socio-environmental and/or heritable risk factors associated with the prevalence of NCDs and NMHs, or others.

Theoretical considerations for the proposal

It has been proposed to describe the practices, to know the evolution and trends of relevant data management, since, from them, in addition to our experience in the development of information systems, good practices will be proposed, so that they represent guidelines for action, that is, guidelines or recommendations based on the evolution, development trends and practical experience.

In the current existing digital ecosystem, there is a separation between the data generated as part of specific purpose applications, in many cases of commercial use for the purposes of the application, with copyrights on data structures, i.e., on data dictionaries, table structure and their relationships (Altman et al., 2007), (Groth et al., 2014). In another aspect, data generated as a result of the publication of academic data (Bourne et al., 1997), (Benson et al., 2013), (Callaghan et al., 2014), (González-Beltrán et al., 2014), (González-Beltrán et al., 2015) and in a last one, the data generated through initiatives under the Creative Commons license, (Berman et al., 2003), (Bauch et al., 2011), (Crosas, 2011), (Velankar et al., 2014), (Starr et al., 2015), (Papoutsoglou et al., 2020). From these three main aspects it can be seen that far from unifying, they cause the circumstance that unfortunately prevents extracting the maximum benefit from investments in the generation and collection of data, especially in health, whatever the purpose from which they were derived (Kjetil-Sandve et al., 2013), (Casino, 2019), (Alterovitz et al., 2020), since depending on the aspect in which the data of a given application is framed or boxed, will be the feasibility of identifiability to make them massively sharable or not.

Exploring and critically analyzing what is related to the identifiability of data, which is what potentially makes them available as an "informational asset", metadata is considered in the literature as an engineering resource, and there is a diversity of works in which obstacles are reflected and highlighted, both technical, behavioral, and governance, that affect the quality of metadata and propose possible solutions, either to mitigate the issue of metadata quality, or even the lack of metadata, especially in public data repositories-resources (Harland, 2012), (Kinjo et al. , 2012), (Lecarpentier et al., 2013), (Martone, 2015), (Musen et al., 2015), (Roche et al., 2015), (Rose et al., 2015), (Szostak, 2019), (Kaliyaperumal et al., 2020).

Certainly, what is not found is any protocol on the important role concerning the government-management, for purposes of preservation, data conservation, present and subsequent statistical collection for its use and exploitation, on populations in general or specific. In fact, if the government is not considered essential for data purposes of a general nature, less specifically in health. Perhaps this is because this could well be considered a management issue, that is, data management, which would lead to move from an engineering role to a data management role (data stewards) (Wenger et al., 2000), (Wolstencroft et al., 2015), (UniProt, 2015), (Ohno-machado et al., 2015), (Susanna-Assunta et al., 2012). Perhaps even, the lack of literature regarding the subject is because such a role is often underappreciated, whose concern is to make traceable statistics of everything, health statistics, we consider that it should be the occupation and concern, in the case of health, of local, regional and/or global health agencies.

Metadata, which are the data about data, the descriptions of the data themselves, are the essential component to enable the collection of data as an asset, for NCDs, NMH or health, they are the core of any data exchange system. It is this metadata that can potentially drive the discovery of unobvious associations and link seemingly unrelated datasets. It is even the metadata itself that provides the essential context, describing with data what the data collector defined, for example, who generated the dataset, in what time period, in what population and how, which is certainly a data management or data governance task that someone must take on.

The important role of curation and stewardship of health or other data should be appreciated, even though it often is not. Generating a collection of metadata, which would identify the collection of databases available for exploration and analysis, would provide visibility of the richness of the data asset, which would provide a basis for critically linking related and unrelated topics. The lack of metadata and/or the quality of this data is not overlooked, hence it is considered that local, regional or global health agencies should take care to generate rules that address the identification of the real asset "data". Such rules should make it possible to overcome technical and behavioral obstacles, as well as those affecting the quality of metadata, in order to collectively contribute to the construction of solutions for the enrichment of health "data" for humanity.

Proposed model

Given that the main requirement is to obtain quantitative data for the deep understanding, prevention and solution of health issues, a problem for which it is proposed to contribute with the generation of guidelines that promote the conservation and management of health data, from the development of Apps that enable the promotion of survey, observational, prospective and longitudinal studies, with total privacy, thus making the handling of sensitive information confidential, this contribution is considered highly valuable.

In this sense, the collection of data, of what represents information or personal data, with respect to what are diagnostic data and results, relevant data that provide guidance on health, must be segregated, so that, in a data analysis (observation), these are not mixed, with the intention of avoiding at all costs that if data are compromised, it is not feasible to disseminate personal data with diagnoses of the particular health of any individual. If the above is achieved, then, what will be feasible is the dissemination, as wide as possible, of the health data of the population or populations in the specific, in order to develop strategies for prevention, diagnosis and treatment of NCDs and NMH in general, or even monitoring them as already proposed by the World Health Organization (World Health Organization, 2019), (World Health Organization, 2020), without particularizing in any person, since it may be considered victimization.

Thus, in response to this generic but crucial requirement, a tripartite data management-governance model is proposed as shown in Figure 1, to contribute data in the absence or insufficiency of such data:

– A first instance, acting as controller, which will be in charge of determining what to present "from the data instance", to the users, because this would be the data and access control. This will enable the segregation of the personal data that would live in this instance, with respect to the data of the observation or observations, which will be deposited at all times in the data instance.

– A second instance of the model, the data instance, which would contain the data obtained through the Apps, such as: variables and constants, as in the case of NCDs and NMH, some clinimetric batteries, specifications, among others.

– A third instance, of visualization or view, in a graphical interface, which would be in charge of showing only what the controlling instance orders.
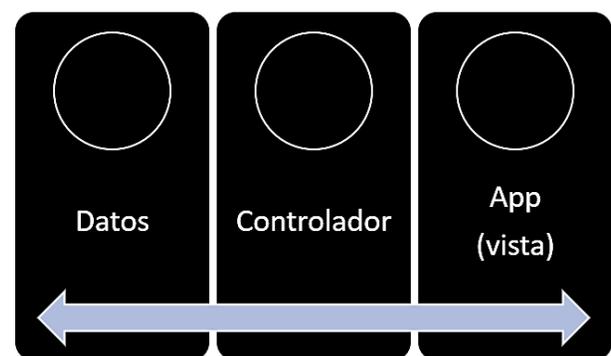

Figure 1 Proposed tripartite governance-management model (to ensure the preservation and management of health data)
Source: Own elaboration

With the proposed model (Figure 1), given the approach of indirect controller-data and controller-view interaction, communication between the data and the view, ideally will never occur and therefore is not contemplated, because the data are independent of the user view-interface, which enables total anonymity, a functionality that is highly desired, although not necessarily materialized, in this type of applications.

Certainly, what the model offers is to articulate, for the purpose of preserving and managing health data, the instances/programs/projects to generate health data via Apps, which will have to take the model as a guideline to follow, and thus harmonize all development efforts through the model. In such a way that the "active data" is stored with total anonymity, and is the one that is preserved over time, so that it is integrated into the available data via the metadata of the data, so that at the required time it can be processed, analyzed and interpreted and information can be extracted.
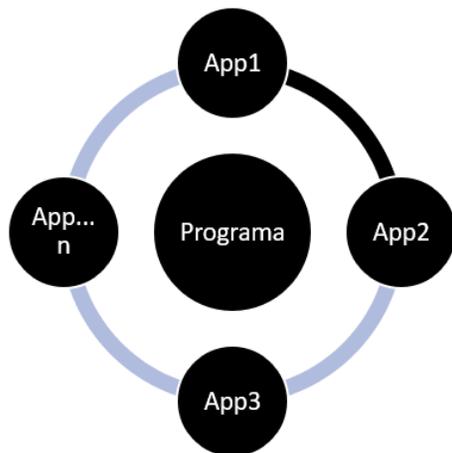

Figure 2 Program or portfolio that, through projects (of Apps), generates affinity - data collection
Source: Own elaboration

The above models (Figure 1 and Figure 2) can feasibly be proposed as framework schemes for structuring the development of applications for use in health, and even, in a future work, as a de facto format guideline for data and metadata in health (only in the logical design part and storage objects, Figure 1), for the purpose of obtaining standardized survey, observational, prospective and longitudinal data, and for the development of orchestrated apps (Figure 2).

Even the same model (Figure 2) can orchestrate programs and projects for the development of databases, in terms that the model can be the structuring scheme (of use), which will give applicative independence, that is, of view or face to the user, and fully structured in the part of the conformation of the databases, which will enable more data, by adding projects that can share data among them, and within the program or portfolio of projects.

Furthermore, the proposed model will ensure that the data generated, regardless of the instance that generated them, will be analyzed and transformed into useful, accessible and understandable information for those interested in the analysis of the databases that can be built, and the phenomena that the data contained therein may describe (complex or not).

The model, given the anonymity scheme of health data, can be used for data collection, allowing the monitoring, evaluation and analysis of the health situation, in relation to a particular disease, a campaign in action or any other health analysis need that is sought to be computed. Similarly, for surveillance, research and control of general and specific risks, in explicit contexts. In addition, it can also provide data to know the social participation in health, imposing the use in general or specific populations, which through invitation or obligation must respond through the Apps. Finally, it can contribute to the development of policies, regulation and compliance with them, as it can constitute a monitoring tool, which through the model guarantees privacy.

Good practices

These practices are written with the intention that they can act as a guideline for those who wish to contribute to obtaining quantitative data for the deep understanding, prevention and solution of health problems, through standardizing the structuring of the approach of the program or specific project, so that from the beginning good data management is considered, and managed data and metadata standardization, for the reuse of the data that can be generated. In this sense, the practices outlined, product of both the evolution and trends of government-data management, as well as our experience in the development of information systems, are the following:

– Propose the development of programs or project portfolios, at the local level (at the level of municipalities), regional (at the state or federal level) or global level, in the medium and long term, that generate data with statistical (historical) trend, with well-defined periodicities. Within the programs, it is imperative to contemplate projects as short term as feasible, such that these projects allow obtaining, as a dividend, data on a medium and regular basis (Figure 2).

– It must be ensured that the information generated in each of the projects participating in the program can be processed and integrated into a local, regional or global database repository, and ensure its accessibility, interoperability, efficient use, visualization, communication and dissemination through standardized metadata.

– For implementation, the data will have to be collected through the use of multiplatform mobile device applications (App), which make use of the proposed model based on the privacy of the patient being studied (Figure 1), whose information is anonymous, confidential, and used exclusively for study and statistical purposes.

– It is expected that the data will allow the identification of correlations, for example, to determine whether there is a relationship between behaviors such as substance use and NCDs in a generational way, in addition to their comorbidity with mental disorders, so as much data as possible should be collected, regardless of whether or not a use has been assigned to them at the outset, through, for example, survey-type clinimetric scales.

– Use general guidelines for the construction of databases, such as those presented in the FAIR Principles Guide (Wilkinson et al., 2016), for data management (Kaliyaperumal et al., 2020), with the relevant dimension for health data, among which the following stand out:

– Findability. That is, the data and the databases where they live must have the appropriate and relevant format, which guarantees that they can be found, located or retrieved.

– Accessibility. This in terms of enabling the right to enter, providing, if necessary, sufficient access credentials and authorizations for such purpose, in terms of guaranteeing access to the databases or data, to whoever must have access.

– Interoperability. For which it is proposed that, within the scheme of government-management of health data, the WHO take the lead for global data, and the federation, states and municipalities do the same for data at the country-region level, so that the standard to be followed for health data is defined. This would ensure that the metadata, data and their repositories comply with the standards defined by the governing body in question, and thus achieve the widest possible interoperability with other databases generated at the local, regional or global levels. In the absence of a global steering committee or one centered on federal bodies, it is feasible to establish local steering committees at the state or municipal level to remedy the general deficiency, but to provide guidelines, albeit at the local level.

– Reusability. In the same case as the previous one, it is necessary to establish guidelines and define the standard to be followed for health data in terms of reusability. Data and their repositories must comply with standards that allow their use for different but related purposes, and even very different from those for which they were originally intended, with the intention of achieving intensive use, i.e., the widest possible use, so as to obtain the greatest benefits from the informational asset: data.

Conclusions

There is a clear need for the integration, processing, analysis and visualization of health data, as well as for local, regional and global statistics, and the development of applications (Apps), which contribute to the generation and collection for the incubation of data banks for use in health, is a viable alternative in the face of this need. It is foreseeable that, through the Apps developed, taking as a guide the good practices proposed here and the proposed model, the study of general or specific populations can be carried out, from the diagnostic approach, with psychiatric instruments (batteries) or others, giving certainty of not violating the privacy of the user who makes use of the App, when putting into practice the "model" at the moment of the implementation of the system, which will keep the anonymity.

It is also predictable that the use of Apps will lead to the generation of massive databases, which in turn will allow searching for possible socio-environmental and/or heritable risk factors, for example, associated with the prevalence of NCDs and NHM, of general or specific populations, through periodic and ordered statistical data, which are not available today.

The fact that the scales, for the example or another, would be raised through the use of applications on smartphones / mobile devices (Apps), multiplatform, whose information will be anonymous and confidential, will encourage the use, since it must be remembered that today many smartphone users give companies like Facebook or Google data and information like never before, over what the same populations are willing to deliver to the governments-managements of different nations or global agencies such as WHO. Given the reluctance to deliver data to the government, and not so to social networking companies or others, despite the fact that they would be used exclusively for public health, academic and statistical purposes, being the Apps the viable and suitable means of capture/diagnosis to promote the collection of data in which there may be opposition from the population, for the communication of these, being part of modern technologies it is expected they can also make the delivery for the purpose of health, by the simple fact of being an attractive technology.

By having data, in bases such as those described, it is to be expected that through such it is possible to determine if there are relationships or correlations between behaviors such as substance use, NCDs and NHM generationally, in addition to their comorbidity with other mental disorders or other indoles.

Undoubtedly, thinking about medium and long term programs, from the integration and analysis of one or several databases, it will be possible to obtain information that will be useful for local, regional and/or global decision making, in addition to providing feedback on the action-reaction effects. The fertilization, through programs with projects that make use of Apps as a tool, with a view to building large health databases, would contribute to support a quantitative overview of the general local, regional and/or global health situation, and/or specific scenarios such as NCDs and NMHs in the same orders.

The democratization and transparency of health data for the generation of useful, accessible and understandable information for the benefit of local, regional or global health is plausible with the use of technological tools such as Apps. But it is unavoidable and necessary to use government-management schemes for data management, to promote the conservation and administration of health data, to promote articulation, to avoid the generation of "data silos" that do not allow integration, the model and the good practices proposed being an approximation.

As a subsequent application work, the proposed model and the described practices will be implemented in a program, with its relevant projects of local or regional scope. In this sense, in order to be able to execute it, we will seek to obtain the pertinent funds that will allow the execution of the program.

One line of work that we can visualize for the WHO is the normative definition and worldwide implementation of government-health data management. This is to circumvent and meet the need for data generation and collection, with this world body as the governing body, in addition to establishing the technical guidelines for metadata, logical design and storage objects, and the guidelines regarding the risks derived from the quality, security and life cycle of health data, which will enhance the use of these in processes, systems and operation of institutions locally, regionally or globally.

References

Alterovitz, G.; Heale, B.; Jones, J.; Kreda, D.; Lin, F.; Liu, L.; Liu, X.; Mandl, K.; Poloway, D.; Ramoni, R.; Wagner, A.; Warner, J. (2020). FHIR Genomics: enabling standardization for precision medicine use cases. *npj Genomic Medicine*, 5(1), 1-4.

Altman, M.; Crabtree, J.; Donakowski, D.; Maynard, M. (2007). Data Preservation Alliance for the Social Sciences: A Model for Collaboration. *Ann Arbor*, 1001, 1-15.

Benson, D.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.; Ostell, J.; Sayers, E. (2013). GenBank. *Nucleic Acids Research*, 41, D36–D42. *https://doi.org/10.1093/nar/gks1195*

Berman, H.; Henrick, K.; Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. Nature Structural and Molecular Biology, 10, 980. *https://doi.org/10.1038/nsb1203-980*

Bauch, A.; Adamczyk, I.; Buczek, P.; Elmer, F. J.; Enimanev, K.; Glyzewski, P.; Kohler, M.; Pylak, T.; Quandt, A.; Ramakrishnan, C.; Beisel, C.; Malmström, L.; Aebersold, R.; Rinn, B. (2011). openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics*, 12, 1-12.

Bourne, P.; Berman, H.; McMahon, B.; Watenpaugh, K.; Westbrook, J.; Fitzgerald, P. (1997). The Macromolecular Crystallographic Information File (mmCIF). *Meth. Enzymol*, 277, 571–590.

Callaghan, S. (2014). Joint Declaration of Data Citation Principles. *FORCE11*.

Casino, G. (2019). Retos y perspectivas en el ecosistema de la información biomédica. *El profesional de la información*, 28(2). *https://doi.org/10.3145/epi.2019.mar.05*

Crosas, M. (2011). The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data. *D-Lib Mag,* 17 (1), 2.

González-Beltrán, A.; Eamonn M.; Susanna-Assunta, S.; Rocca-Serra, P. (2014). linkedISA: semantic representation of ISA-Tab experimental metadata. *BMC Bioinformatics,* 15(S4).

González-Beltrán, A.; Li, P.; Zhao, J.; Avila-Garcia, M. S.; Roos, M.; Thompson, M.; van der Horst, E.; Kaliyaperumal, R.; Luo, R.; Tin-Lap L., Tak-wah, L.; Edmunds, S.; Susanna-Assunta, S.; Rocca-Serra, P. (2015). From Peer-Reviewed to Peer-Reproduced in Scholarly Publishing: The Complementary Roles of Data Models and Workflows in Bioinformatics. *PLoS ONE,* 10(7), 1-20.

Groth, P.; Loizou, A.; Gray, A.; Goble, C.; Harland, L.; Pettifer, S. (2014). API-centric Linked Data integration: The Open PHACTS Discovery Platform case study. *Web Semant. Sci. Serv. Agents World Wide Web,* 29, 12-18.

Harland, L. (2012). Open PHACTS: A Semantic Knowledge Infrastructure for Public and Commercial Drug Discovery Research. *Knowledge Engineering and Knowledge Management Lect. Notes Comput. Sci.*, 1–7.

Jacobsen, A.; Kaliyaperumal, R.; Bonino S. S., Luiz O.; Mons, B.; Schultes, E.; Roos, M.; Thompson, M. (2020). A generic workflow for the data FAIRification process. *Data Intelligence*, 56-65.

Kinjo, A.; Suzuki, H.; Yamashita, R.; Ikegawa, Y.; Kudou, T.; Igarashi, R.; Kengaku, Y.; Cho, H.; Standley, D.; Nakagawa, A.; Nakamura, H. (2012). Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. *Nucleic Acids Research,* 40(D1), D453–D460.

Kjetil-Sandve, G.; Nekrutenko, A.; Taylor, J.; Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. *PLoS Computational Biology,* 9(10), 1-4.

Lecarpentier, D.; Wittenburg, P.; Elbers, W.; Michelini, A.; Kanso, R.; Coveney, P.; Baxter, R. (2013). EUDAT: A New Cross-Disciplinary Data Infrastructure for Science. *The International Journal of Digital Curation,* 8(1), 279–287.

Martone, M. (2015). FORCE11: Building the Future for Research Communications and e-Scholarship. *Bioscience,* 65(7), 635.

Musen, M.; Bean, C.; Kei-Hoi, C.; Dumontier, M.; Durante, K.; Gevaert, O.; Gonzalez-Beltran, A.; Khatri, P.; Kleinstein, S.; Connor, M.; Pouliot, Y.; Rocca-Serra, P.; Susanna-Assunta, S.; Wiser, J.; CEDAR team (2015). The center for expanded data annotation and retrieval. *Journal of the American Medical Informatics Association,* 22(6), 1148–1152. *https://doi.org/10.1093/jamia/ocv048*

Roche, D.; Kruuk, L.; Lanfear, R.; Binning, S. (2015). Public Data Archiving in Ecology and Evolution: How Well Are We Doing?. *PLOS Biology,* 13, 1-12. *https://doi.org/*10.1371/journal.pbio.100229

Rose, P.; Prli ́, A.; Bi, C.; Bluhm, W.; Christie, C.; Dutta, S.; Kramer-Green, R.; Goodsell, D.; Westbrook, J.; Woo, J.; Young, J.; Zardecki, C.; Berman, H.; Bourne, P.; Burley, S. (2015). The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Research,* 43, D345–D356. *https://doi.org/*10.1093/nar/gku1214

Starr, J.; Castro, E.; Crosas, M.; Dumontier, M.; Downs, R.; Duerr, R.; Haak, L.; Haendel, M.; Herman, I.; Hodson, S.; Hourclé, J.; Ernest-Kratz, J.; Lin1, J.; Holm-Nielsen, L.; Nurnberger, A.; Proell, S.; Rauber, A.; Sacchi, S.; Smith, A.; Taylor, M.; Clark, T. (2015). Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science,* 1.

Susanna-Assunta, S.; Rocca-Serra, P.; Field, D.; Maguire, E.; Taylor, C.; Hofmann, O.; Fang, H.; Neumann, S.; Tong, W.; Amaral-Zettler, L.; Begley, K.; Booth, T.; Bougueleret, L.; Burns, G.; Chapman, B.; Clark, T.; Lee-Ann C.; Copeland, J.; Das, S.; de-Daruvar, A.; de-Matos, P.; Dix, I.; Edmunds, S.; T-Evelo, C.; Forster, M.; Gaudet, P.; Gilbert, J.; Goble, C.; Griffin, J.; Jacob, D.; Kleinjans, J.; Harland, L.; Haug, K.; Hermjakob, H.; Ho-Sui, S.; Laederach, A.; Liang, S.; Marshall, S.; McGrath, A.; Merrill, E.; Reilly, D.; Roux, M.; Shamu, C.; Shang, C.; Steinbeck, C.; Trefethen, A.; Williams-Jones, B.; Wolstencroft, K.; Xenarios, I.; Hide, W. (2012). Toward interoperable bioscience data. *Nature Genetics,* 44(2), 121-126.

Szostak, R. (2019). A synthetic approach to the classification of music. Review article. *El profesional de la información*, 29(1). *https://doi.org/*10.3145/epi.2020.ene.05

The UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Research,* 43, D204–D212. *https://doi.org/*10.1093/nar/gku989

O'Dea (2020). Smartphone penetration worldwide as share of global population 2016-2020. *Technology and Telecommunications*, 1 marzo. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

Ohno-machado, L.; Alter, G.; Fore, I.; Martone, M.; Susanna-Assunta S.; Xu, H. (2015). bioCADDIE white paper—Data Discovery Index. *Scientific data,* 1-30. http://doi.org/10.6084/m9.figshare.1362572

Papoutsoglou, E.; Faria, D.; Arend, D.; Arnaud, E.; Athanasiadis, L.; Chaves, I.; Coppens, F.; Cornut, G.; Costa, B.; Ćwiek-Kupczyńska, H.; Droesbeke, B.; Finkers, R.; Gruden, K.; Junker, A.; King, G.; Krajewski, P.; Lange, M.; Laporte, M. A.; Michotey, C.; Oppermann, M.; Ostler, R.; Poorter, H.; Ramírez-Gonzalez, R.; Ramšak, Ž.; Reif, J.; Rocca-Serra, P.; Sansone, S. A.; Scholz, U.; Tardieu, F.; Uauy, C.; Usadel, B.; Visser, R.; Weise, S.; Kersey, P.; Miguel, C.; Adam-Blondon, A. F.; Pommier, C. (2020), Enabling reusability of plant phenomic datasets with MIAPPE. *New Phytologist*, 1-43. *https://doi.org/*10.1111/nph.16544

Statista (2020). Global smartphone penetration rate as share of population from 2016 to 2020. *Technology and Telecommunications*, March 1, *https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/*

Velankar, S.; Alhroub, Y.; ̈lle-Alili, A.; Best, C.; Boutselakis, H.; Caboche, S.; Conroy, M.; Dana, J.; van-Ginkel, G.; Golovin, A.; Gore, S.; Gutmanas, A.; Haslam, P.; Hirshberg, M.; Melford, J.; Lagerstedt, I.; Mir, S.; Newman, L.; Oldfield, T.; Penkett, C.; Pineda-Castillo, J.; Rinaldi, L.; Sahni, G.; Sawka, G.; Sen, S.; Slowley, R.; Sousa S., Alan W.; Suarez-Uruena, A.; Swaminathan, J.; Symmons, M.; Vranken, W.; Wainwrightand-Gerard M. (2014). PDBe: Protein Data Bank in Europe. *Nucleic Acids Research,* 39, D402–D410. *https://doi:10.1093/nar/gkq985*

Wenger, M.; Ochsenbein; E.; Dubois; B.; Borde; G; Jasniewicz; Lalo ̈e; L.; Monier (2000). The SIMBAD astronomical database-The CDS reference database for astronomical objects. *Astronomy Astrophysics Supplement Series,* 143, 9–22.

Wilkinson, M.; Dumontier; A. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data,* 3, 1-9. *https://doi.org/10.1038/sdata.2016.18*

Wolstencroft, K.; Owen, S.; Krebs, O.; Nguyen, Q.; Stanford, N.; Golebiewski, M.; Weidemann, A.; Bittkowski, M.; An, L.; Shockley, D.; Snoep, J.; Mueller, W.; Goble, C. (2015). SEEK: a systems biology data and model management platform. *BMC Systems Biology,* 9(33), 1-12. *https://doi.org/10.1186/s12918-015-0174-y*

World Health Organization, (2019). Global Health Observatory (GHO) data, 1 de diciembre de 2019. *https://www.who.int/gho/ncd/en/*

World Health Organization, (2020). Noncommunicable diseases and mental health, March 1, 2020. *https://www.who.int/nmh/about/en/.*