Oral Oncology 98 (2019) 8–12

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Review

The potential use of big data in oncology

Stefan M. Willems[a,b,*], Sanne Abeln[c], K. Anton Feenstra[c], Remco de Bree[d], Egge F. van der Poel[e], Robert J. Baatenburg de Jong[e], Jaap Heringa[c], Michiel W.M. van den Brekel[f]

[a] Department of Pathology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
[b] Department of Pathology, Netherlands Cancer Institute, Amsterdam, the Netherlands
[c] Department of Computer Science, Faculty of Science, Vrije Universiteit, Amsterdam, the Netherlands
[d] Department of Head and Neck Surgical Oncology, University Medical Center Utrecht, Utrecht, the Netherlands
[e] Department of Head and Neck Surgery, Erasmus Cancer Center, Erasmus MC, Rotterdam, the Netherlands
[f] Department of Head and Neck Oncology and Surgery, Netherlands Cancer Institute, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Big data
Personalized medicine
Head and neck cancer
FAIR data

ABSTRACT

In this era of information technology, big data analysis is entering biomedical sciences. But what is big data, where do they come from and what can we do with it? In this commentary, the main sources of big data are explained, especially in (head and neck) oncology. It also touches upon the need to integrate various sources of clinical, pathological and quality-of-life data. It discusses some initiatives in linking of such datasets on a nationwide scale in the Netherlands. Finally, it touches upon important issues regarding governance, FAIRness of data and the need to bring into place the necessary infrastructures needed to fully exploit the full potential of big data sets in head and neck cancer.

Introduction

Big data and the computer technology to analyze it are called one of the top 10 revolutions in the coming decade [1]. It is foreseen that its impact parallels that of the Internet, the cloud, and, more recently, block-chains (known from crypto-currencies as the bitcoin) [2]. Big data phenomena are penetrating in virtually all sectors. On large scale, they have been first applied by information power companies (IBM, Google, Facebook, Amazon). Algorithms, using neural networks and machine-learning techniques have been developed and are used by these large IT-oriented companies to predict behavior of people and use this information for person-oriented marketing. Also health insurance companies and governments have large interest in big data developments and big data have entered life sciences too. But what is big data and what can we do with it? [3].

Definition of big data

Though many people and companies use the word “big data”, they may not always mean the same, or interpret it in the same way. Most of us have a vague notion of what it could be (“anything that won’t fit an excel sheet”), but big data is not just synonymous to “a lot of data”. A way to define big data in health care, is its description according to the 5 V’s (https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data The 5 Vs of Big data, September 17, 2016 Anil Jain). From this definition, big data contain:

- Volume: big data are of big size, containing a lot of data points/records of multiple subjects. These include diagnostic work-up [clinical, radiological, pathological), treatment data (surgery, systemic therapy, radiotherapy and their combinations), response data and complications.
- Velocity: big data has two velocity aspects: [1] big data are created at an increasingly high speed, and [2] they have to be computed/digested relatively fast. Worldwide the incidence of cancer is increasing, while patients live longer. Together with the technological advances and monitoring devices, an increasing number of data will have to be processed within the same time.
- Variety: big data comprise a huge variability of data types. This variety has important opportunities (many different data types enrich the quality and usefulness of it), and challenges regarding its heterogeneity warranting standardization (synoptic reporting).
- Variability: it’s crucial to realize that data capturing varies in place and time. Capturing a (predefined) mandatory minimal dataset is a prerequisite to get most of (synoptic) reported data. This doesn’t only need consensus on the minimal data required; it also involves

* Corresponding author at: Department of pathology, University Medical Center Utrecht, Utrecht University, The Netherlands, Heidelberglaan 100, 3581 CX, Utrecht, the Netherlands.
E-mail address: s.m.willems-4@umcutrecht.nl (S.M. Willems).

https://doi.org/10.1016/j.oraloncology.2019.09.003
Received 5 March 2019; Received in revised form 31 July 2019; Accepted 6 September 2019
Available online 12 September 2019

univocal definitions (e.g. recurrence vs residual disease).
- Value: setting up a data infrastructure to collect and interpret data is only worthwhile, when it enables generation of data-derived conclusions or measurements based on accurate data that can really lead to measurable improvements or impact in health care.

While the sheer size of data collections is often an issue, an even more pressing problem is that data resources are typically spatially distributed across the globe and deposited in ways that make it difficult to integrate the data. This may lead to scalability problems since the data need to be transported via the internet, but, more importantly, requires harmonization and standardization efforts if the data is to be integrated and used in a common workflow to answer an overarching query.

Why do we want big Data?

Big data are practical useful in various areas. Location tracking helps logistic companies to mitigate risks in transport, speed and reliability of delivery. In the financial world, Securities Exchange Commission (SEC) uses big data network analytics to identify possible frauds. Entertainment companies such as Netflix and YouTube use past views and online behavior to increase engagement and drive more revenues. Advertising companies are probably the biggest big data players. Data analyzed from Facebook, Twitter and Google monitor behavior and transactions that advertisers use to run targeted campaigns. Breeding companies use drones flying over the crop fields, sending back imaging data to inform the breeding process. With big data, hospitals can improve monitoring of intensive care patients. Efficiency of (expensive) medication can be measured and epidemic outbreaks can be forecasted in an early stage.

More specifically for the medical disciplines, big data could be helpful in developing and reshaping disease prevention strategies. Combining large data sets of genomics and environmental data will help to predict which individuals/groups are at risk for developing certain (chronic) diseases and cancer. This might elicit specific actions aimed to influence the environmental factors and behavior that contribute to health risks in target groups. Big data will also be helpful to evaluate current prevention programs and might help to identify novel insights to improve these. Also in a therapeutic setting, big data are instrumental to monitor e.g. the effects of specific therapies, such as those of expensive oncolytics, especially in relation to patient and tumor (genetic) characteristics. This will help to improve precision medicine and fuel important knowledge to calculate cost-efficiency of certain treatment regiments.

Data sources of Big Data in Medicine

Sources for big data are plentiful and can be of different kinds. In oncology, the most obvious are patient derived data. These include various data points/subject and are usually recorded in electronic patient files for clinical purposes. These data contain the clinical data of patients, tumors, treatment and outcome and may include demographic details such as gender and age, presenting symptoms, family history, comorbidity, radiological data (such as CT, MRI, PET, US) as well as solid and liquid tissue-based analysis (such as histopathological diagnosis/features, immunohistochemistry, DNA/RNA sequencing experiments, blood analyses and whole genome BAM files which contains on average ~ 100 GB). But also data from *in vitro* experiments are important and can be an important source. A second source of big data includes the computational analysis of these data. These processed data comprise indirect and computed data, including radiomics and digital image analysis as well as genetic expression and mutation analyses. An increasing source of these processed data come from machine learning and usually contains large computer data files of structured data.

A third source of big data comes from the patients themselves, including patient related outcome measurements (PROMs) and patient related experience measurements (PREMs) who record all kinds of measures using apps on computers and mobile devices, either provided by their caregivers (eHealth, telemedicine) or by their own initiative.

A fourth source is from published literature (IBM project). As every year over 1 M biomedical articles are published, there is no doctor in the world who can read even just a fraction of the knowledge published, let alone all relevant textbooks and other internet sources.

Nevertheless, one key factor stands out for big data in oncology: the depth (volume) of data per patient. In oncology many observables (thousands to even millions) per patient are routinely generated and stored, while typical cohort sizes of patients are relatively small. This imbalance in the depth of data per patient versus the cohort size is even more prominent for rare cancer types such as head and neck cancer. However, recent methodological advances in machine learning and neural network are specifically powerful if there are instances to learn over. For example, object recognition in images works very well, but thousands to millions of examples are needed to optimize such methods. Hence, if we want to translate this to optimize personalized treatments, we also need data depth in the number of samples [4]. This makes good data keeping, data standardization, data sharing, data provenance and data exchange protocols essential for oncology, and absolutely necessary in the field of head and neck oncology.

Integration of big data in head and neck cancer (HNC)

So as sources are multiple and volumes are high, standardization in data capturing is warranted. This standardization will lead to much more uniform and complete datasets, which are easier to link to other data resources. Standardization is essential for data integration, which is needed for data interpretation and creating value from the data. E.g. to monitor the quality of care after specific surgical interventions, it's crucial to understand the case mix of patients in terms of tumor stage, (neo) adjuvant therapies, co-morbidity, etc.

In The Netherlands, several national databases have been established to structurally capture clinical, pathological, genetic/genomic

Table 1
Data sources for most common cancer types in The Netherlands, including HNC.

Tumortype	Clinical	Pathological	Genetic/genomic	Radiological	PROM/PREM
HNSCC	DHNA	PALGA	PALGA/HMF		NET-QUIBIC
breast cancer	NBCA	PALGA	PALGA/HMF		
lung cancer	DLCA/NVALT	PALGA	PALGA/HMF		
prostate cancer		PALGA	PALGA/HMF		
CRC	DSCA	PALGA	PALGA/HMF		
melanoma	DMTR	PALGA	PALGA/HMF		

For most common tumor types, collections of various data (clinical, pathological genomic/genetic) have been well organized, with the exception of radiological data.
DHNA: Dutch Head and Neck Audit; NBCA: National Breast Cancer Audit; DLCA: Dutch Lung Cancer Audit.
DSCA: Dutch Surgical Colorectal Audit; DMTR: Dutch Melanoma Treatment Registry; PALGA: Pathologisch Anatomisch Landelijk Gegevens Archief; HMF: Hartwig Medical Foundation.

data and PROM/PREMs (Table 1). Clinical data are being reported since 2014 in the Dutch Head and Neck Audit (DHNA), which was incorporated in the Dutch Institute for Clinical Auditing (DICA) 2017 who installed subgroups for specific disease types (cancer and non-cancer). Tumor tissue based data such as pathological and genomic/genetic data have been structurally collected nation-wide, and now also synoptically reported for > 20 different tumor types. In contrast, structured data basing of radiological data is still lacking. In addition to other cancer types, for head and neck, the NET-QUBIC consortium has initiated the national platform to report PREMs/PROMs for HNC. So for HNC in The Netherlands, most patient-derived data sources have now been installed. Internationally, many similar initiatives are underway, e.g. the head and neck squamous cell carcinoma (HNSCC) collection [5] and The Cancer Imaging Archive (TCIA) [6]. The essential next step is to bring these data together now, preferentially on the individual patient-level.

In the Netherlands, head and neck cancer care is centralized in 8 head and neck centers with 6 preferred partners and united within the Dutch Head and Neck Society (NWHHT). This places the head and neck cancer community in the ideal position to unite the expertise and endeavors to optimize uniform data input and roll out of the current separate databases. Moreover, head and neck cancer is in the ideal position to establish nationwide linkage of these databases and develop algorithms for integrated data-analysis.

Utility of big data

The future potential of big data (in biomedical research) is not fully clear yet. For today (and tomorrow), big data will create value for [1] daily diagnostics, [2] quality of care/life (including PROMs and PREMs) and [3] biomedical research [7]. We will give some examples of currently available applications.

Daily diagnostics

Big data can already have relevance in every day clinical practice. An example is the near-real time access Dutch pathologists have to the nationwide histopathological follow up of each individual patient. The PALGA foundation governs all digital histopathological records in The Netherlands ever since 1971 (www.palga.nl). Containing over 72 million records of over 12 million patients in The Netherlands, PALGA is one of the largest biomedical databases in the world and covers all 55 pathology labs in The Netherlands. Every time a Dutch pathologist authorizes a histopathology report, one copy is stored in the local hospital information system, and one copy in the central PALGA database. So, this database contains real time pathological follow up of each patient that is directly visible for each PALGA member (pathologist or molecular biologist). This offers huge potential in recognizing relevant patient (oncological) history, e.g. ruling out a recent malignancy in cases of a suspect tumor of unknown primary; or offering pathological documentation on previous relevant pathological features (such as resection margins and positive lymph node) in case pathology was performed in another lab. Also co-occurrence of diseases or unknown associations in low prevalent disease, that at first sight seem not to be related can be studied using this database [8].

Electronic patient files generate an enormous amount of medical data, which can be used for prognostic modeling. One of the first prognostic models for HNC patients receiving care at medical centers in developed countries is available online at www.oncologiq.nl [9]. Automatization of statistical prognostication processes allows automatic updating of models when new data is gathered [10]. These data can also be used to develop clinical decision making tools for improved patient counseling and non-binary patient related outcome measurements.

Quality of care measurements

Linking databases on patient outcomes with data on patient characteristics and treatment can offer unprecedented potential for feeding back quality an efficacy of care. Recently, a French study showed the landscape of molecular testing for targeted therapy in non-small cell lung cancer (NSCLC) in France and subsequent treatment regiments based on this [11]. This allows direct feedback on optimal test-treatment correlations. More importantly, it might be a strong incentive for underperforming labs, to revise their protocol/workflow to improve their optimum of care. Also in The Netherlands, linking data from the national cancer registry (containing clinical stage, treatment and outcome data) with the aforementioned PALGA database, has been able the show the variety in clinical care in head and neck cancer in The Netherlands [12,13]. Though improving the quality of care can only be reached by transparency on such data, it should be realized that feedback of such data, especially outcome data and when benchmarked, can only be done with indisputable prudency as labs and hospitals might fear reputation damage or naming-and-shaming [14]. In practice, when published anonymously to the public and fed back disclosed only on the individual level, experience learns that most hospitals are actually happy to cooperate in such mirror feedback. This has led in The Netherlands to the development of algorithms for automatic feedback of pathology and treatment related items on a regular basis, such as the Dutch Institute for Clinical Auditing (www.dica.nl). Mirror information showing higher recurrence rates than those in peer hospitals, possible will be an incentive to zoom in on the underlying chain to identify (and solve) potential weaknesses.

Biomedical research

Probably most benefit will be generated from big data in the field of research. The leading era of “genome wide association studies” (GWAS) has been broadening towards an era of “data wide association studies” (DWAS), with a central place for big data. Increase of data, both due to increased used of imaging and molecular analyses and combinations with other data, offer a matchless Walhalla for each data scientist and bioinformatician. Big data fill an unmet need in biomedical research. For example, an important limitation of today’s medicine is our poor understanding of the biology of disease. Only by aggregating huge amounts of big data, all relevant multisource variables, such as DNA, RNA, protein and metabolomics data will aggregate and can be integrated in more realistic models to predict how tumors will behave and which patients will benefit best form specific therapies. These integrated multi-omics data will for example provide more comprehensive insight into biological behavior and mechanisms that underlie growth patter, metastatic potential as well as response to (targeted) treatment of HNSCC.

Personalized medicine

From the perspective of turning our current understanding and available data into actionable insights that can be used to improve treatment outcome, personalized medicine is absolutely dependent on big data [15]. The amount of data available for the biomedical community exponentially increases, especially with advancing technologies generating terabytes of data, notably in sequencing and imaging. In terms of quantity, most data do not come from direct, patient related records available from daily clinical practice, but to a larger extent from computed automatic data analyses such as radiomics and digital image analysis. Head and neck cancers present a unique set of diagnostic and therapeutic challenges by nature of its complex anatomy and heterogeneity. Radiomics holds the potential to address these barriers [16]. Radiomics extracts and mines a large number of medical imaging features in a non-invasive and cost-effective way. The underlying assumption of radiomics is that these imaging features quantify

phenotypic characteristics of an entire tumor. Radiomics in precision oncology and cancer care allow for prognostic and reliable machine-learning methods for the stratification (or personalization), i.e. identifying differences in (expected/predicted) survival between (groups of) patients, and prediction of treatment outcome(s) to support selection of the best possible treatment of head and neck cancer patients [17]. This might enable medical and radiation oncologist to (de-)escalate systemic treatment and irradiation doses in specific patient populations.

FAIR data

To ensure data can be reused in secondary studies, it is essential they adhere to the FAIR (Findable, Accessible, Interoperable, Reusable) principles. These FAIR data principles have been first published in 2014 [18]. Since then, the principles have been recognized and endorsed by the G20 (2016) and the G7 (2017), while the EU has taken FAIR data at the heart of the European Open Science Cloud (EOSC). An important part of FAIRness of a data resource is concerned with its metadata, where findability (F) is reliant on having a persistent identifier in place; accessibility (A) requires clearly defined access rules (data privacy constraints are within the definition) and licensing; and interoperability (I) is dependent upon employing a community-recognized ontology for describing the data. Finally, provenance of the data and accurateness and completeness of the *meta*-data is essential for the reusability (R) of the data.

Challenges and future perspectives

Though already applied in current clinical practice, and with tremendous promises ahead, producers of big data also face challenges to make them optimally useful in life sciences. First of all, with rapidly developing technologies, such as next generation sequencing (especially whole exome/genome sequencing) and radiomics, the volume of data continues to increase exponentially. These huge amounts of data add an increasing complexity that might impede data interpretation. This holds especially true when the increase in data (velocity and volume) is also paralleled by an increase in the heterogeneity of data (variability), including treatments, outcomes, differences in study design, analytical methods and interpretation pipelines, which hamper drawing firm conclusions from the data.

A second challenge comes with the proper governance of data, especially when linked from various sources. How and which data are made available, who is the owner of the data? Does the patient still have governance of this own datasets? Or does the researcher governs it, and if so, which researcher, or the treating physician, the data generator, or is it the person who tries to make sense out of the data (e.g. the computational biologist or medical bioinformatician)?

Repositories and databases for archiving and sharing biomolecular patient data

In a (bio)medical research setting, the aim is often to obtain as many data as possible from as many patients and hospitals as possible, while privacy issues as well as security and protection measures (GDPR) often prohibit the availability.

One key challenge is to store patient identifiable data, such as genome sequences, in such a way that the data can be reused for other studies, while safe guarding the privacy of the patients from which the data was collected (www.phgfoundation.org). Whereas for large data sets in other domains, such as computer science, open accessibility may be preferred, here privacy concerns must clearly outweigh a desire for complete openness. The European Genome Archive (EGA) is a purpose-built database to store raw sequencing data. For each study, with data stored in EGA, there is a strong role for a data access committee (DAC), which is governed by the research initiative that collected the data and can decide to provide access to the data upon request by other researcher [19]. A secondary challenge is posed, when researcher would like to browse processed biomolecular data, without the ability to trace back individual markers. Here several solution are available that summarise the data without showing individual markers [20], or other repositories solve this by fine graining access control [21]. Lastly, a remaining challenge is to link the different data resources in a privacy aware manner, while being able to track the exact computational processing that has been performed on the data; several initiatives have made the first steps to achieve such linking [22,23,4].

The potential use of big data in life sciences and head and neck oncology is tremendous. It might also transform the way we share clinical and research data. Instead of individuals or organizations physically sharing datasets, the (near) real time/streaming of data together with the huge volume of data, will make it impossible to keep exchanging data sets like we do today. Instead of bringing together all kind of datasets in a central comprehensive database, a likely scenario might be that big data users will develop more organic, decentralized virtual networks, such as envisioned in the personal health train by the Dutch Techcentre for Life Sciences (DTL) [24]. Within these networks, databases are connected as nodes, accessible under predefined conditions to users. Increased connectivity and (thus) complexity will also demand new ways of interpreting data, as well as translating these data and its interpretations back to the individual patient. This latter requires big data-derived knowledge computed from all these data sets to be translated to the specific “small data” environment of the individual care dependent patient. For this last step, crucially also integrating intuitive and emotional aspects, we still need medical professionals. Bed side manners are for the foreseeable future well out of reach of big data or machine learning approaches.

Conclusion

The value of big data capturing relies on the volume, velocity variety, veracity of various, often complex, data sets. Integration of these sources is key and will be beneficial for improvements in biomedical research, patient care and monitoring quality of care. In The Netherlands, where head and neck cancer care is centralized and various national big data resources are in place, there is an unique opportunity to unite, link and integrate these data and fulfill this unmet need. Such a head and neck cancer infrastructure should optimize data input as well as (bioinformatical) data integration including FAIRification.

Funding

This work was supported by the Dutch Cancer Foundation (KWF-project number UU-2017-8225 to SMW).

Declaration of Competing Interest

None declared.

Acknowledgements

The authors thank Martine D’Herdt for carefully reading and commenting the manuscript.

References

[1] Shaikh AR, Butte AJ, Schully SD, et al. Collaborative biomedicine in the age of big data: the case of cancer. J Med Internet Res 2014;16(4):e101. https://doi.org/10.2196/jmir.2496. Apr 7.
[2] Roman-Belmonte JM, De la Corte-Rodriguez H, Rodriguez-Merchan EC. How blockchain technology can change medicine. Postgrad Med 2018;130(4):420–7.
[3] Bourne PE. What Big Data means to me. J Am Med Inform Assoc 2014;21(2):194. https://doi.org/10.1136/amiajnl-2014-002651.
[4] Zhang C, Bijlard J, Staiger C, Scollen S, et al. Systematically linking tranSMART,

galaxy and EGA for reusing human translational research data. F1000Res. 2017;6. https://doi.org/10.12688/f1000research.12168.1. Aug 16 ELIXIR-1488.
[5] Grossberg AJ, Mohamed ASR, El Halawani H, et al. Sci. Data 2018;5:180173. https://doi.org/10.1038/sdata.2018.173. Sep 4.
[6] Prior F, Smith K, Sharma A, et al. The public cancer radiology imaging collections of The Cancer Imaging Archive. Sci Data. 2017;19(4):170124https://doi.org/10.1038/sdata.2017.124. Sep.
[7] Bousfield D, McEntyre J, Velankar S, et al. Patterns of database citation in articles and patents indicate long-term scientific and industry value of biological data resources. F1000Research 2016;5(160). https://doi.org/10.12688/f1000research.7911.1.
[8] Ooft ML, van Ipenburg J, Braunius WW, et al. A nation-wide epidemiological study on the risk of developing second malignancies in patients with different histological subtypes of nasopharyngeal carcinoma. Oral Oncol 2016;56:40–6.
[9] Datema FR, Ferrier MB, Vergouwe Y, et al. Update and external validation of a head and neck cancer prognostic model. Head Neck 2013;35(9):1232–7.
[10] Datema FR, Moya A, Krause P, et al. Novel head and neck cancer survival analysis approach: random survival forests versus Cox proportional hazards regression. Head Neck 2012;34(1):50–8.
[11] Barlesi F, Mazieres J, Merlio JP, et al. Routine molecular profiling of patients with advanced non-small-cell lung cancer: results of a 1-year nationwide programme of the French Cooperative Thoracic Intergroup (IFCT). Lancet 2016;387(10026):1415–26.
[12] Petersen JF, Timmermans AJ, van Dijk BAC. Trends in treatment, incidence and survival of hypopharynx cancer: a 20-year population-based study in the Netherlands. Eur Arch Otorhinolaryngol 2018;275(1):181–9.
[13] Timmermans AJ, van Dijk BA, Overbeek LI, et al. Trends in treatment and survival for advanced laryngeal cancer: A 20-year population-based study in The Netherlands. Head Neck 2016;38(Suppl 1):E1247–55.
[14] de Ridder M, Balm AJ, Smeele LE, et al. An epidemiological evaluation of salivary gland cancer in the Netherlands (1989–2010). Cancer Epidemiol 2015;39(1):14–20. Feb.
[15] Govers TM, Rovers MM, Brands MT, et al. Integrated prediction and decision models are valuable in informing personalized decision making. J Clin Epidemiol 2018. Aug 28 pii: S0895-4356(18)30447-5.
[16] Wong AJ, Kanwar A, Mohamed AS. Radiomics in head and neck cancer: from exploration to application. Transl Cancer Res 2016;5(4):371–82.
[17] Parmar C, Grossmann P, Rietveld D, et al. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Front Oncol 2015;3(5):272.
[18] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;15(3):160018https://doi.org/10.1038/sdata.2016.18.
[19] Lappalainen I, Almeida-King J, Kumanduri V, et al. The European genome-phenome archive of human data consented for biomedical research. Nat Genet. 2015;47(7):692–5.
[20] Klonowska K, Czubak K, Wojciechowska M, et al. Oncogenomic portals for the visualization and analysis of genome-wide cancer data. Oncotarget 2016;7(1):176–92. Jan 5.
[21] Christoph J, Knell C, Bosserhoff A, et al. Usability and suitability of the omics-integrating analysis platform tranSMART for translational research and education. Appl Clin Inform. 2017;8(4):1173–83.
[22] He S, Yong M, Matthews PM, et al. TranSMART-XNAT connector tranSMART-XNAT connector-image selection based on clinical phenotypes and genetic profiles. Bioinformatics 2017;33(5):787–8. Mar 1.
[23] Hoogstrate Y, Zhang C, Senf A, et al. Integration of EGA secure data access into galaxy. F1000Res 2016;5. https://doi.org/10.12688/f1000research.10221.1. Dec 12pii: ELIXIR-2841. eCollection 2016.
[24] Eijssen L, Evelo C, Kok R, et al. The Dutch techcentre for life sciences: enabling data-intensive life science research in the Netherlands. F1000Research 2015. https://doi.org/10.12688/f1000research.6009.2.