

# Connected Patients in Light of Big Data

THE JOURNAL 2022

Diogo Neves, Henrique Martins  
Linking Patients With Data: Brain-Computer  
Interfaces and Healthcare Innovation

Eugene Fidelis Soh et al.  
Flipping Healthcare Through a Population  
Health Stack

Srdjan Babic et al.  
Data Science in Modern Healthcare

Francisco Maestre et al.  
DICOM Metadata - A Useful Resource for  
Big Data Analytics

Jörg Schwarz  
How to Utilise the Massive Amount of  
Health Data Collected by Consumers to  
Improve Health Outcomes

Betsabe Melcon et al.  
In Search of Gold in Health Data





# DICOM Metadata - A Useful Resource for Big Data Analytics

**Francisco Maestre** | CTO & Big Data | Idonia Medical Exchange | Palma de Mallorca, Spain

**Ana Paula Sorrell Meriño** | Biomedical Engineer | Researcher | Spain

**Christian Mata Miquel** | Scientific Researcher and Project Manager | Fundació de Recerca Sant Joan de Déu | Associate Professor | Universitat Politècnica de Catalunya (UPC) | Spain

**Miguel Cabrer** | Founder and CEO | Idonia Medical Exchange | Palma de Mallorca, Spain

This article provides an overview of new ways to represent data combining patient access and DICOM information, advanced use of medical imaging metadata, analysis of radiation dose and image segmentation and deep learning for feature engineering to enrich data.

## Introduction

Data is the world's [most valuable resource](#) and it is possible to find data everywhere. In medical images, data covers not only gigapixel images, but also metadata and quantitative measurements (Aiello et al. 2021). [DICOM](#) (Digital Imaging and Communications in Medicine) is a clear source of medical data, since it is the current standard for storing and transmitting medical images (Aiello et al. 2021) and related information (Savaris et al. 2014); this means it contains raw data imaging and all metadata related to the procedures of image acquisition and curation (Aiello et al. 2021).

Idonia is a medical imaging exchange platform that facilitates the collection, storage, delivery and visualisation of medical images for medical centres, professionals and patients. Over 140 million medical images (DICOM) have been processed and delivered up to today. But with that, a necessity arose: studying the medical data contained in them in order to see if any conclusions could be drawn from such examination. This was possible thanks to the Idonia Magic Link, a tool to deliver medical images to patients that replaces the use of CD/USB, which simplified the process of accessing the data.

## Data Lake for Medical Imaging Activity

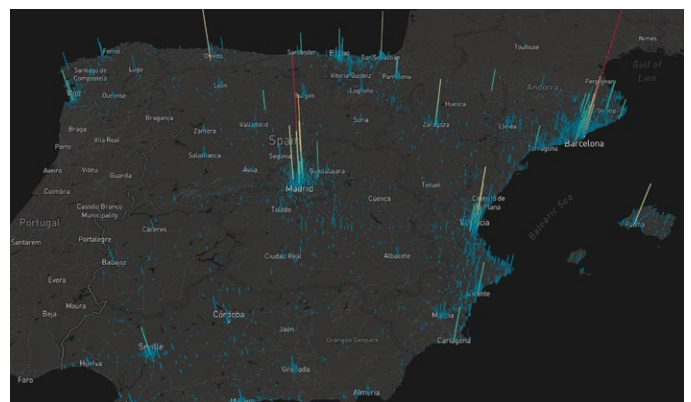
After two years of research, Idonia created the Idonia Data Lake under an R&D project supported by the CDTI (*Centro para el Desarrollo Tecnológico Industrial*) from the Ministry of Science and Innovation of Spain. The purpose of this tool was to analyse all the information provided by the clients –hospitals at most– to provide them with relevant and new information about their own data.

The first analysis was done around medical imaging activity

combined with some information stored in DICOM images, such as the imaging technique, manufacturer of the medical device, the patient's body part examined and modality of study, etc. This information was shared with clients via the development of a command centre functionality that went beyond traditional dashboards; it also contained access information from both professionals and patients.

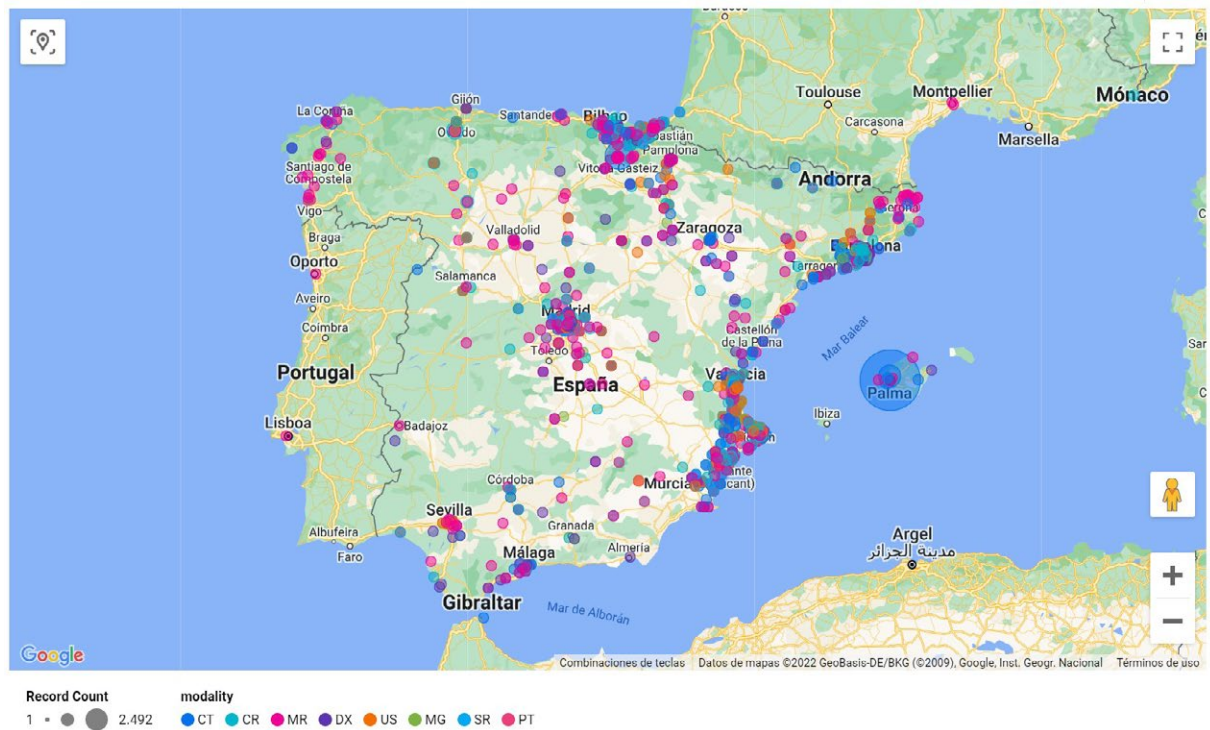
The picture below shows an example of information from patients accessing their medical studies through Idonia Magic Link.

It was also possible to generate a medical device activity map, that provided some useful information about medical devices, generated information, their activity and remote access to their generated content.



## Data Analytics Around Radiation

After years of studying medical related information, it was found that the most relevant data relied on radiation dose



parameters, and from this, the idea of developing a digital tool to analyse radiation-related data originated.

In the current defined Directives and Regulations, no limits on the radiation dose are stipulated for patients undergoing diagnostic or treatment procedures. There is proof that ionising radiation has direct implications in human health (ICRP 1990), which is why measures need to be taken as soon as possible.

These actions start with being able to quantify the radiation received by a patient in studies over time, which can be done by means of a DICOM dataset of metadata. With the available DICOM data, it is possible to quantify radiation by the estimation of the effective dose, a parameter that serves as a dose descriptor that contemplates the biological sensitivity of the irradiated tissue and reflects the risk of a non-uniform whole-body exposure (AAPM Report 2008).

The purpose of the research project (Sorrell et al. 2022) was to extract and analyse the data and compare it with the SEPR (*Sociedad Española de Protección Radiológica*) and the ICRP (*International Commission on Radiological Protection*) recommendations. The problem with DICOM is that it is only a standard, which means it is not mandatory to implement and its use is pretty open. The result of this is differently structured/organised data sets, which make data analysis a complex challenge. Since the information is valuable, different techniques can be applied in order to collect that information.

### Radiation-Related Information

The imparted ionising radiation to a patient has always been the main concern in radiology since there is substantial evidence of adverse effects due to radiation exposure. *Image*

*Gently* and *Image Wisely* are two initiatives that tend to optimise the act of radiating a patient to make sure that the radiation dose is only the necessary – and never more than what's strictly necessary – to produce images with a good enough quality for the diagnosis (Sorrell et al. 2022).

The process of ionisation that occurs when imparting radiation to a patient necessarily changes atoms and molecules and may thus sometimes damage cells. This may result in preventing the cell from surviving or reproducing, or subsisting as a modified cell; both outcomes have profoundly different implications for the organism as a whole (ICRP 1990).

Historically, the quantities used to measure the amount of ionising radiation have been based on the gross number of ionising events in a defined situation or on the gross amount of energy deposited, usually in a defined mass of material. The *Absorbed Dose (D)*, for example, is defined as the energy absorbed per unit mass, and the units used for such measurement are Grays (Gy) (ICRP 1990). The problem with this is that it does not take into account the biological effects of the tissue when exposed to radiation, and this is something to always have in mind when studying radiation dose.

The *Equivalent Dose (H)* is the absorbed dose averaged over a tissue or organ (rather than at a point) and weighted for the radiation quality that is of interest. This has to be done because the probability of stochastic effects depends not only on the absorbed dose but also on the type and energy of the dose. Even though the equivalent dose is only multiplied by a no-units factor, the units to express it are different from the absorbed dose; now it is given in Sieverts (Sv) (ICRP 1990).

The [European Council Directive 2013/59/EURATOM](#), of 5



December 2013, establishes uniform basic safety standards for the protection of the health of individuals subject to occupational, medical and public exposures against the dangers arising from ionising radiation. It defines medical exposure as the exposure incurred by patients in order to be diagnosed or treated of any disease. In disparity with the dose limits for professional workers and the public in both occupational and public exposures, neither the Directive nor any of the Spanish Regulations establishes any limits on patient dose. In fact, in [Article 6.1](#), it says that the radiological protection of the exposed patient will be optimised in order to keep the individual doses *as low as reasonably possible* (ALARP). [Article 6.2](#) follows up with this and establishes that, in medical exposures, dose restrictions will only apply with respect to the protection of caregivers and volunteers involved in medical or biomedical research.

### CT Scanners

Let's take [Computed Tomography](#) (CT) as an example. CT scans consist of a computerised x-ray imaging procedure in which a narrow beam of x-rays is aimed at a patient and quickly rotates around the body.

CT images are based on the different x-ray absorption rates of the various organs of the human body, which is why it provides both good soft tissue resolution (contrast) as well as high spatial resolution (Zhanli et al. 2009). To ensure the best resolution, the dose imparted must be considerably high; in fact, the dose levels imparted in CT exceed those from conventional radiography and fluoroscopy and the use of CT continues to grow, often by 10% to 15% per year, which leads to a discussion of radiation risk versus medical benefit (AAPM Reports 2008). At the end of the day, what matters are the long-term repercussions of [radiation exposure](#), which is why the American Association of Physicists in Medicine (AAPM) has defined several dose parameters to provide guidance on reasonable CT dose levels on routine examinations (AAPM 2008).

focused on [defining statistical](#) averages for population dosimetric studies and Diagnostic Reference Levels (DRL) to optimise radiological procedures.

DRLs are defined in Article 4 of the EU Directive as dose levels in medical radiodiagnostic or interventional radiology practices, among others (Sorrell et al. 2022). They use a dose index magnitude in order to quantify the ionising radiation used to obtain a medical image. In CT images, DRLs are the CT Dose Index (CTDI) and the Dose Length Product (DLP) (AAPM 2008).

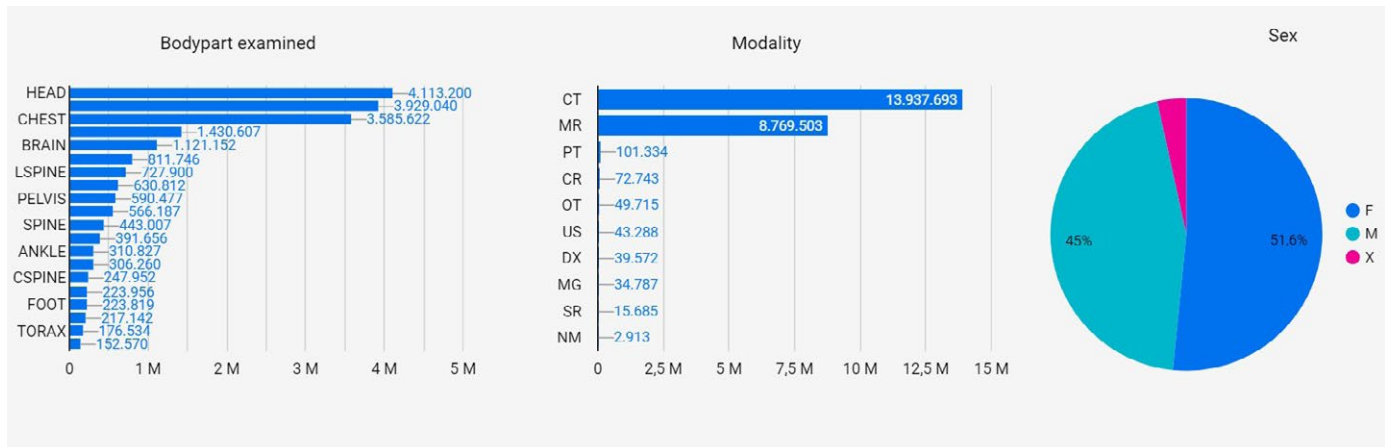
### DICOM Standard

DICOM is the current standard for storing and transmitting medical images (Aiello et al. 2021) and it is defined as the international standard for medical images and related information. It was originally developed by the American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) (Savaris et al. 2014). Its first publication, in 1993, revolutionised the practice of radiology, allowing the replacement of x-ray film with a fully digital workflow (Aiello et al. 2021).

The DICOM standard comprises a set of specifications regarding structure, format, and exchange protocols for digital-based medical images. In other words, it defines the formats for medical images that can be exchanged with the data and quality necessary for clinical use (Aiello et al. 2021). This means it contains raw data imaging and all metadata related to the procedures of image acquisition and curation (AAPM 2008).

DICOM images consist of textual metadata (Kathiravelu et al. 2021); in fact, a DICOM file contains both the image and a large variety of data in the header (Aiello et al. 2021). Physically, the content of a DICOM file can be seen as structured at the data element level (Savaris et al. 2021), which means that the information recorded on the file are the attributes; these shall be ordered by increasing data element tag number and shall occur at most once in a data set (Aiello et al. 2021).

Idonia Data Lake already obtains and analyses some relevant information from DICOM metadata.



### Radiation on CT Scanners

In radiophysics, it is complicated to determine the dose received by a patient, individually, which is why they have

### Radiation Data Stored in DICOM Images and SR Documents

DICOM metadata also includes diagnostic reference levels



such as the CTDI and in some cases, the DLP, which could be used to estimate the Effective Dose. After a large inspection of over 1.5M anonymised CT images, it was found that out of the vast amount of information available, the necessary information to extract out of DICOM images for the estimation was not as broad (Sorrell et al. 2022). For a complete radiation dose-related study attributes found on the DICOM SR are also necessary.

The Effective Dose can be estimated with the formulas defined by the AAPM in their 2008 journal (AAPM 2008), which means the standardised data that matters is:

- either with the body part Examined, the CTDIvol –volumetric CTDI– and the scan length; or,
- the body part examined and the DLP.

As it turns out, not all the information is stored in the same place, and as mentioned previously, it is not always available. Apart from the DICOM image, healthcare centres may send Structured Report (SR) documents, where a lot of the radiation-related data is stored, including the Scan Length and the DLP. Since this information is not always sent, it is not accessible in most cases, which leads to the development of several alternatives to estimate the effective dose. Some of these studies include artificial intelligence algorithms for image segmentation –in order to define the scan length– and body part recognition –in case the body part examined is unknown (Juszczak et al. 2021).

## Conclusion

The delivery of medical images in a secure and efficient way is a necessary functionality up to date, and especially, patients are taking benefit from it. Accumulating those delivered images in a cloud infrastructure, and analysing the data generated around them may bring great benefits for medical centres to better understand their patient profiling and medical activity. An infinite loop of value is created.

DICOM files contain both the image and a large variety of

metadata. This metadata provides valuable information for many different applications, including a radiation-related study. At the presented project, a very extensive radiation dose study was performed by means of the DICOM metadata. It was done in collaboration with some Idonia clients with the idea of analysing a vast anonymised DICOM images dataset to see if any radiation dose-related conclusions could be drawn from such examination. This entailed a large inspection of the DICOM standard in order to be able to identify the data relevant for the study; only then it would have been possible to estimate the effective dose for an ionising radiation risk control.

The data exploration executed was remarkable throughout and allowed a deep understanding on how DICOM files are constructed. The value of aggregating all the information from different sources in one common dataset (Data lake) opens new possibilities to analyse or enrich the data that can benefit all data providers. Image segmentation and deep learning can be used for feature engineering techniques in order to enrich data when it's not available over the base of common data set information.

DICOM has some very interesting medical metadata available, thus it can complement a Big Data and Analytics project in the medicine scope. The Idonia medical imaging delivery service based on cloud technology allows to combine disparate information like the remote accessibility from patients and professionals with the DICOM metadata. This unique way to analyse the information is just a first step but the potential is huge. The aggregate information obtained from different sources through cloud infrastructure not only enriches the data source but provides more potential capabilities for deep learning data analytics.

## Conflict of Interest

None. ■

## REFERENCES

Aiello M, Esposito G, Pagliari G et al. (2021) How does DICOM support big data management? Investigating its use in medical imaging community. *Insights into Imaging*. 12(1):1-21.

American Association of Physicists in Medicine [2008] The Measurement, Reporting, and Management of Radiation Dose in CT. Available from [https://www.aapm.org/pubs/reports/rpt\\_96.pdf](https://www.aapm.org/pubs/reports/rpt_96.pdf)

ICRP [1990] ICRP Publication 60: Recommendations of the International Commission on Radiological Protection. Available from [https://journals.sagepub.com/doi/pdf/10.1177/ANIB\\_21\\_1-3](https://journals.sagepub.com/doi/pdf/10.1177/ANIB_21_1-3)

Juszczak J et al. (2021) Automated size-specific dose estimates using deep learning image processing. *Medical Image Analysis*. 68:101898.

Kathiravelu P, Sharma A, Sharma P (2021) Understanding Scanner Utilization with Real-Time

DICOM Metadata Extraction. *IEEE Access*. 9:10621–10633.

Savaris A, Härder T, von Wangenheim A (2014) Evaluating a row-store data model for full-content DICOM management. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*. 193–198.

Sorrell AP, Mata C (2022) Analysis of Radiation Dose Using DICOM Metadata. *EEBE*. Universitat Politècnica de Catalunya (UPC). Bachelor's Degree in Biomedical Engineering. Defended 28/06/2022. Barcelona, Spain

Zhanli H, Hairong Z, Jianbao G, Ying Z (2009) Real-time gray and coordinate statistics methods of medical CT image. *3rd International Conference on Bioinformatics and Biomedical Engineering*. 1-4.



**HealthManagement.org**

*Promoting Management and Leadership*