Computational Science Alliance

New Technologies For Storing And Transferring Personal Data

N. Shchegoleva, D. of Sc., Professor,
N. Zalutskaya, Cand. of Sc. (Med), **Assoc**. **Prof**.
A. Dambaeva, B., **J. Kiyamov**, PhD student,
B. A. Dik, PhD student

Saint Petersburg State University,
Federal state budgetary institution «Bekhterev National Medical Research Psychiatry and Neurology Center»,
State Marine Technical University,
St. Petersburg, Russia

INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND ITS APPLICATIONS ICCSA 2022

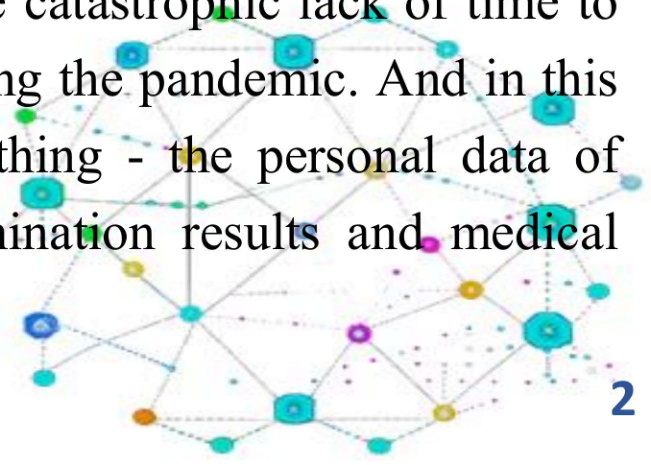
THE INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND ITS APPLICATIONS

Problem

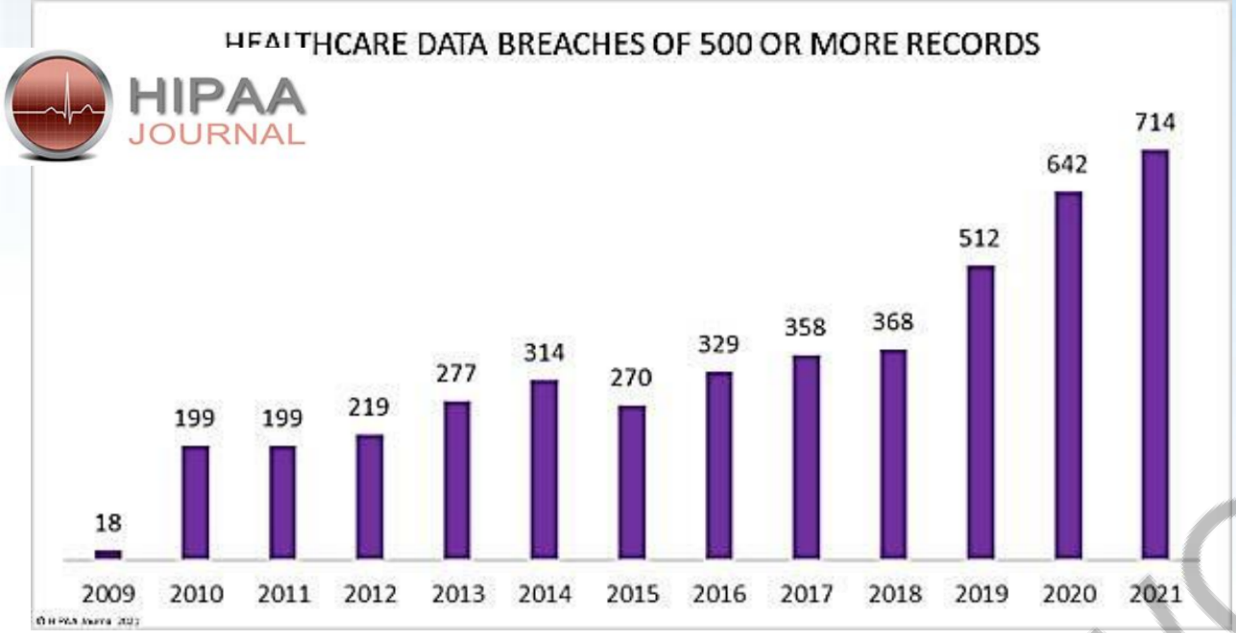
According to the Data Leakage & Breach Intelligence (DLBI) intelligence service, **medical institutions are among the organizations in which data breaches are most often recorded**.

Varonis, an information security company, believes that **the losses of medical institutions** in 2020 from **data leaks amounted to $7.13 million, an increase of 10,5%** compared to the previous year, and judging by the dynamics of 2021, these losses will increase. This is evidenced, for example, by the appearance in the public domain of data of 300,000 people who recovered from COVID-19, the publication of stolen Pfizer data on the COVID-19 vaccine on the network, etc.

“The life cycle” of a leak in the healthcare sector - from the commission of an attack to its detection and remediation - **reached a record 329 days in 2020**." Obviously, this is due to the catastrophic lack of time to analyze incidents in the health sector, all efforts of which were aimed at combating the pandemic. And in this situation, health care institutions were unable to adequately protect the main thing - the personal data of citizens, including information protected by law on the state of health, examination results and medical recommendations.



Introduction



Cybersecurity experts note **new trends** in the **black market**

Information and medical products – various kinds of filing cabinets – are gaining more and more popularity.

It is worth noting a **special category** of medical personal data related to the field of **mental health**.

According to experts of Federal state institution «Bekhterev National Medical Research Psychiatry and Neurology Center»

labeling a **person as mentally ill** often **entails serious social consequences**:

attributing certain negative characteristics to a person, and as a result **falling into social isolation** and **distancing others**, and the **infringement of rights**.

Especially when this information falls into the hands of **incompetent people** who do not have a psychological or psychiatric education, knowledge and experience, but who want to protect other people from a non-existent danger, thus **harming an innocent person**.

Data Encryption

To protect data in practice, several groups of methods are used.

Recently, one of the **most effective approaches is data protection in depth**, that is, the use of several successive measures to prevent third parties from accessing information. This approach uses multiple levels of data protection, which increases the security of the system and reduces the risk of unauthorized access to information.

We propose to **use two levels of protection**.

The **first** is implemented using a **cryptographic encryption** algorithm that provides data protection by converting it into a form that is incomprehensible and useless for an attacker.

The **second** level of protection is provided with the help of **steganography**, that is, an algorithm that hides not only the content of the stored information, but also the very fact of storing or transmitting some data embedded in the image. Since a person "at a glance" is not able to accurately estimate the difference between the size of an image without data embedded in it and with data, the main task is to make the embedded data invisible.

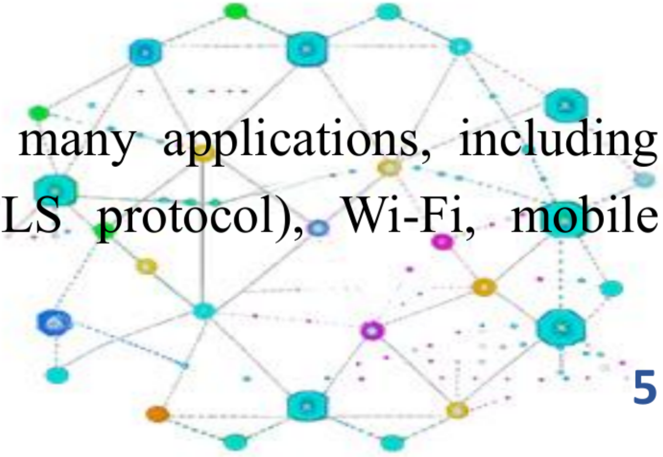
Encryption Algorithm

An analysis of the proposed encryption technologies showed that the **most promising** is the use of **symmetric encryption technology.** In this case, the block cipher is used to encrypt the identifier using a secret key, which is both a pseudo anonymization secret and a recovery secret.

The **advantage** of symmetric encryption algorithms is their **simplicity**, since one key is used for both encryption and decryption, so such algorithms are much faster than their asymmetric encryption counterparts, require less computing power, **do not reduce Internet speed, and allow you to quickly encrypt a large the amount of data**.

The **most promising** and one of the most common symmetric encryption algorithms **is AES** (advanced encryption system). It was developed as an alternative to DES and became the new encryption standard after being approved by NIST in 2001.

Today, AES is the most popular encryption algorithm – it is used in many applications, including security: wireless applications, processors and files, websites (in SSL/TLS protocol), Wi-Fi, mobile applications, VPN (virtual private network), etc.



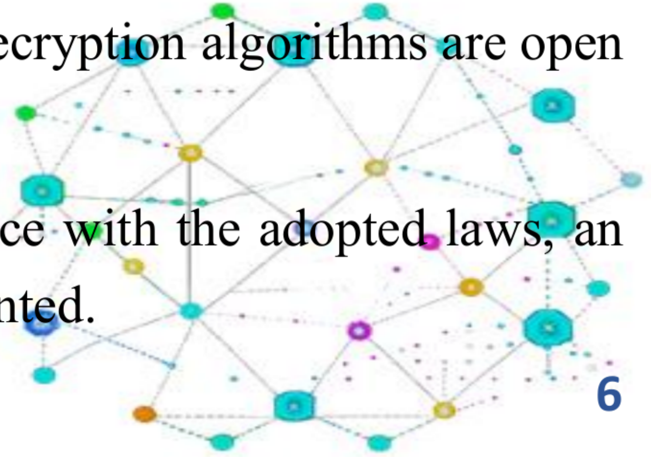
Use of Barcodes

To store and analyze the results of various tests and medical examinations of patients at the **V. M. Bekhterev Research Institute** an information system was developed should **ensure the protection of personal data**, as well as **research results** containing information that **allows you to identify the patient directly or indirectly**.

Currently, **bar codes** are widely used in most medical institutions for **fast and accurate management of patient flows**. However, when realizing access to information, financial and material resources within the framework of a corporate network through barcodes in a significant number of medical institutions, there is an **almost open transmission of patient information** through communication channels.

In this case, there are many opportunities for intercepting confidential and personal data, their substitution or replacement of these data by modifying them, since the barcode decryption algorithms are open and the programs implementing this process are also in the public domain.

Therefore, in order **to protect the personal data of patients** in accordance with the adopted laws, an additional step of hiding data from being read by illegal users should be implemented.



Algorithm for embedding personal information in a color image

1. Preparation of messages with documentary information.

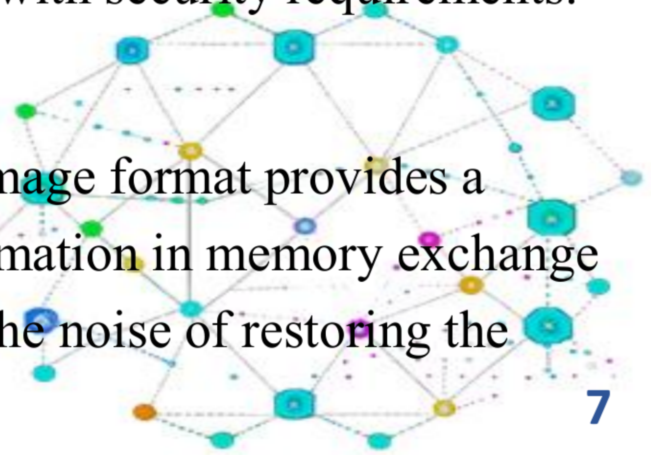
To form a message that will be used to generate a QR-code, it is necessary to group the initial personal textual and digital information (full name, year and place of birth, etc.), biometric data (weight, height, etc.), examination results, etc. by arranging them into semantic groups to record them in three barcodes. At the same time, separator characters (for example, #) can be used to structure information.

2. Encryption of messages with documentary information

Prepared messages with documented information are encrypted using the AES algorithm and key generation, the length of which is chosen by the user in accordance with security requirements.

3. Writing encrypted messages to QR-codes

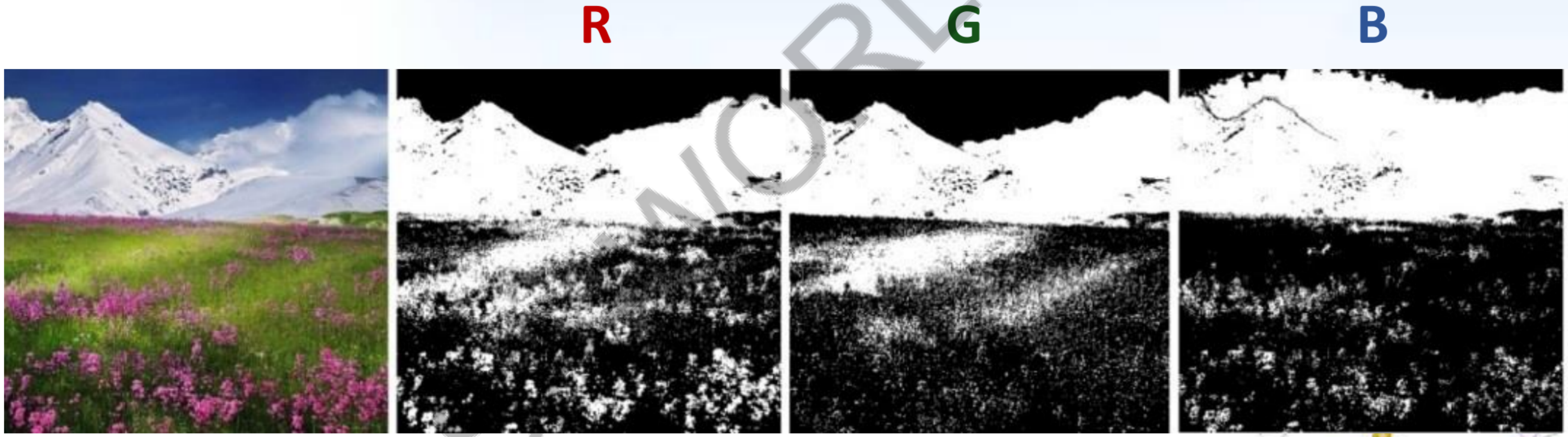
Graphically, QR-codes are halftone images in ".png" format. This image format provides a representation without compression noise, full preservation of information in memory exchange operations, which means that the QR-code will not be distorted by the noise of restoring the recorded information.



Algorithm for embedding personal information in a color image

4. Preparing the container image

The container can be any color image, the size of which exceeds the size of the barcode. To do this, the input color image is decomposed into three components R, G and B, which are three matrices of size M×N (where M and N are the size of the image in pixels).



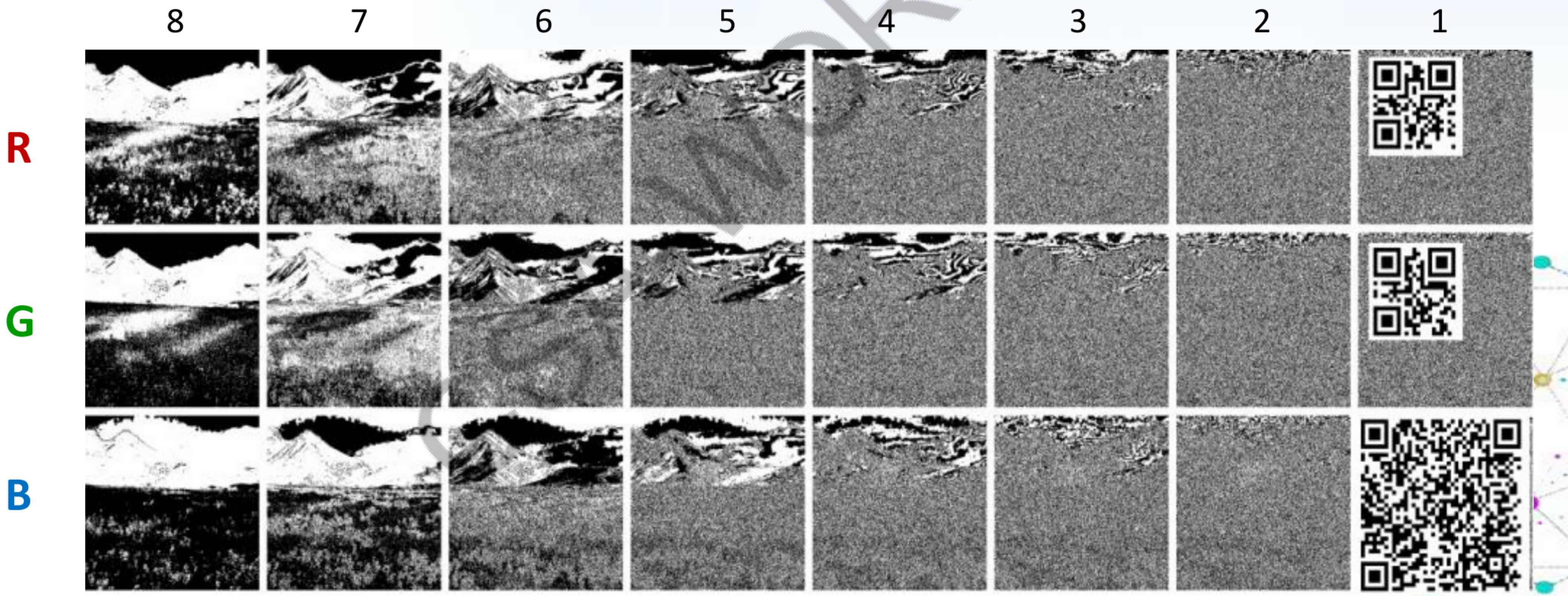
From halftone images, the R, G and B components are removed in the LSB (least significant) layer, the part whose size corresponds to the size of the embedded QR-code.

Algorithm for embedding personal information in a color image

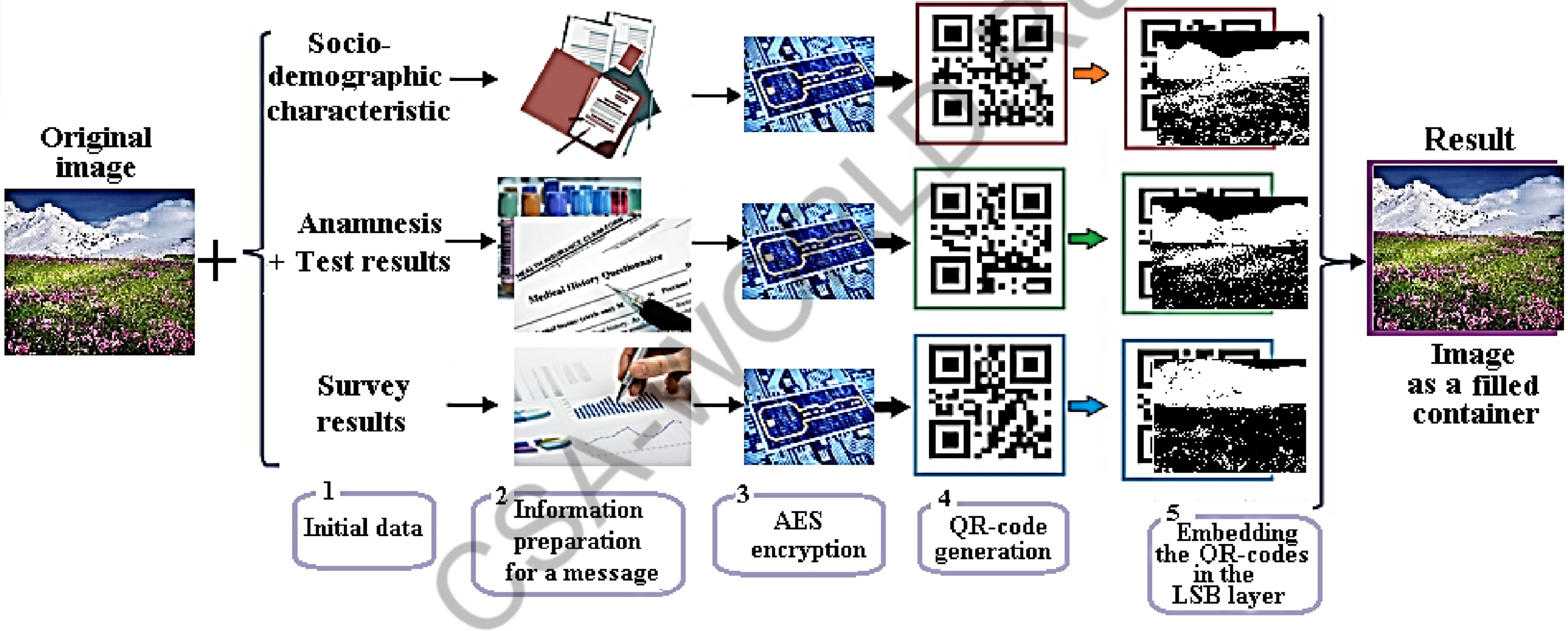
5. Formation of the container image

In place of the removed part of the LSB layer of halftone images, the R, G, B components and the contents of the QR-codes are embedded. Next, the reverse operation is performed – combining all the layers of grayscale images and the R, G and B components into one color image that contains personal information in the form of QR-codes.

R, G, B components of the "filled container image"



Scheme of the process of embedding information in color images



Original image

+

Socio-demographic characteristic

Anamnesis + Test results

Survey results

Result

Image as a filled container

1 Initial data

2 Information preparation for a message

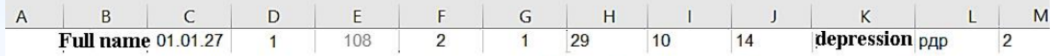
3 AES encryption

4 QR-code generation

5 Embedding the QR-codes in the LSB layer

Practical implementation

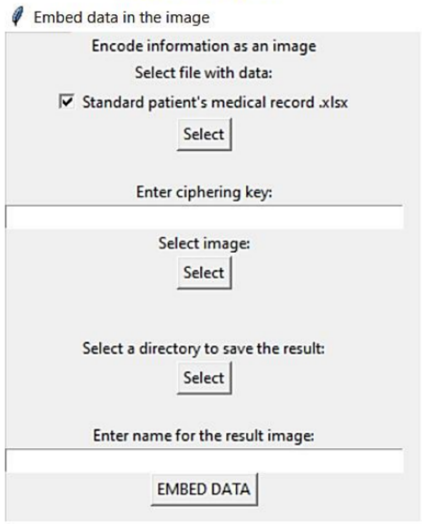
Personal Information



A	B	C	D	E	F	G	H	I	J	K	L	M
	Full name	01.01.27	1	108	2	1	29	10	14	**depression**	рдр	2

Data input for the formation of an image-container

Start window



Embed data in the image

Encode information as an image

Select file with data:

☑ Standard patient's medical record .xlsx

Select

Enter ciphering key:

Select image:

Select

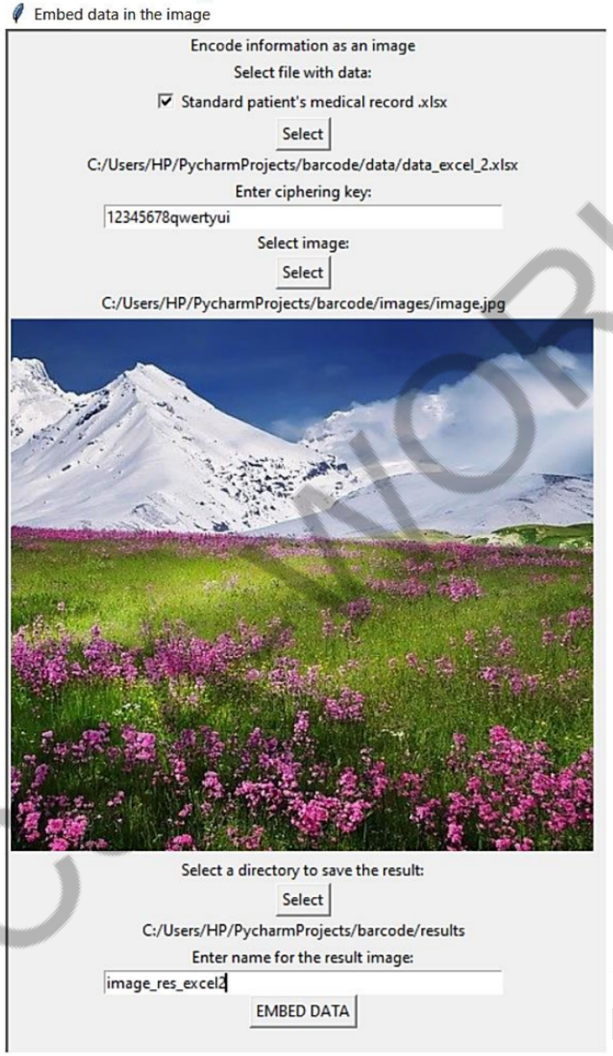
Select a directory to save the result:

Select

Enter name for the result image:

EMBED DATA

Entering data to form a container



Embed data in the image

Encode information as an image

Select file with data:

☑ Standard patient's medical record .xlsx

Select

C:/Users/HP/PycharmProjects/barcode/data/data_excel_2.xlsx

Enter ciphering key:

12345678qwertyui

Select image:

Select

C:/Users/HP/PycharmProjects/barcode/images/image.jpg

Select a directory to save the result:

Select

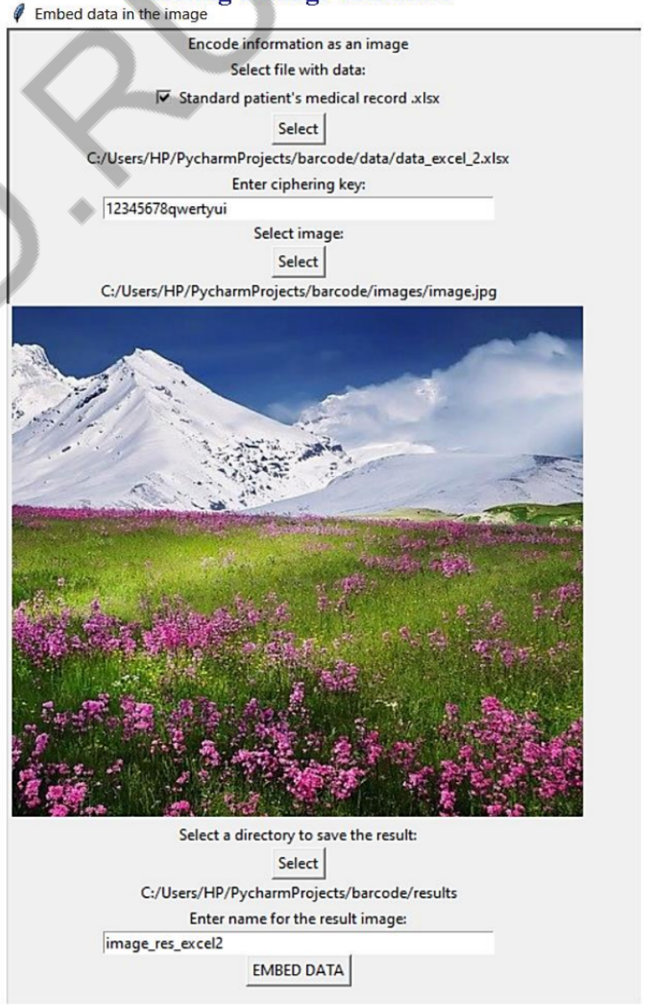
C:/Users/HP/PycharmProjects/barcode/results

Enter name for the result image:

image_res_excel2

EMBED DATA

Selecting a image-container



Embed data in the image

Encode information as an image

Select file with data:

☑ Standard patient's medical record .xlsx

Select

C:/Users/HP/PycharmProjects/barcode/data/data_excel_2.xlsx

Enter ciphering key:

12345678qwertyui

Select image:

Select

C:/Users/HP/PycharmProjects/barcode/images/image.jpg

Select a directory to save the result:

Select

C:/Users/HP/PycharmProjects/barcode/results

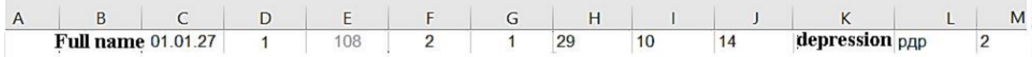
Enter name for the result image:

image_res_excel2

EMBED DATA

The process of extracting data from an image-container

The result of extracting data from the image-container



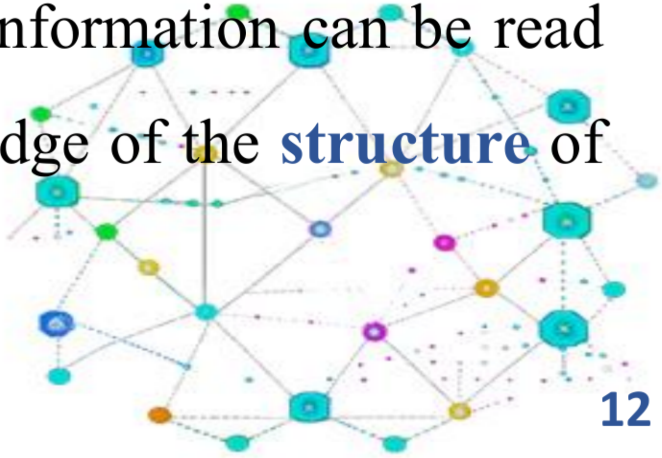
A	B	C	D	E	F	G	H	I	J	K	L	M
	Full name	01.01.27	1	108	2	1	29	10	14	**depression**	рдр	2

Conclusion

It should be noted that **container images** obtained in this way **do not make it possible to detect** the **existence** of documentary information in them.

The use of color images with QR-codes embedded in them containing encrypted information as a storage and carrier of personal data will **ensure the protection** of this information,

as well as **restrict open access** to it, since the embedded information can be read and decrypted **only if the encryption key is present** and knowledge of the **structure** of information is known.



CSA Computational Science Alliance

Thank you for your attention!