Basics of Da(2022-23 – Proge3o “Or(Scolas(ci” (8 CFU)

Application domain – Specifications

We want to create a database to support the citizen science initiative aimed at schools "From botany to big data". The initiative aims to build a support network for schools participating in school garden projects. For each school we want to memorize the name of the school, the mechanographic code, the province, the education cycle (first or second education cycle) and whether or not the institution benefits from funding to participate in the initiative, in in this case we memorize the type.

For each school there is at least one contact person for the initiative, but there may be several. For each person involved we want to store name, surname, email address, optionally a telephone contact and the role (manager, digital animator, teacher, ...). If the school holds funding to participate in the initiative (e.g. funding for the PON EduGreen project) we want to memorize whether the person is the contact person and a participant in the project from which the funding derives. Within the school, there may be multiple classes participating in the initiative. For each of them you want to memorize the class (e.g. 4E), the level (e.g. primary, lower secondary) or the type of school (e.g. applied sciences high school, agricultural) and the reference teacher for participation of this class.

Each school has one or more gardens, identified by a name that identifies the garden within the school. Each vegetable garden can be in the open field or in pots, and is characterized by GPS coordinates and a surface area in m2. We also want to memorize whether the conditions of the garden make it suitable to act as a control for other institutes (i.e. if it is located in a "clean" environmental context and the institute is willing to collaborate with others).

The plants are planted for biomonitoring or phytoremediation purposes. Biomonitoring refers to the monitoring of pollution using living organisms. The main biomonitoring techniques consist in the use of bioaccumulator organisms to provide information on the environmental situation. It provides estimates of the combined effects of multiple pollutants on living beings, has limited management costs and allows covering large areas and diversified territories, allowing for adequate mapping of the territory. Phytoremediation refers to the use of plants to depollute air, water, sediments and soils.

A certain number of species are considered (see Annex 1, which also shows the information to be memorized for each species) for the different purposes and for each species a certain number of replicates (i.e. actual specimens of the plants) are used. In particular, in the case of biomonitoring, the replications of the control group ("in clean conditions") must be the same number as those of the group for which we want to monitor environmental stress. The control replicates may be located in a garden available to the same institute or in a garden made available by another institute and the connection between the group for which environmental stress is monitored and the corresponding control group must be maintained. Specifically, each school should focus on three species and each group should contain 20 replicates.

For each specific plant planted, the species, replica number, group, garden, specific exposure, planting date and the class that planted it will be stored.

The surveys (observations) are carried out on specific plants (replicas) and the information acquired (according to the sheets in Attachment 2) stored with date and time of the survey, date and time of insertion, person responsible for the survey (can be an individual or a class) and responsible for the insertion (if different from the one for the survey and also in this case it can be an individual or a class).

The environmental information relating to pH, humidity and temperature is acquired using sensors or Arduino cards (see Attachment 3, from which the information to be monitored for the different types of sensor/card can be deduced), we want to memorize the number and type of sensors present in each garden (and the replicas associated with that sensor). Information can be

detected via the app and inserted into the database or be transmitted directly from Arduino boards to the database. You want to keep track of how information is acquired.

Tasks to be done

The project consists of the following activities (for each of them the material to be delivered is detailed in the following section).

THE. **Logical design:** The purpose of this activity is the definition of a logical schema relating to a database relational for the domain described above, assuming a workload starting from the specifications provided above (e.g. insertion of schools, vegetable gardens, replications, measurements, data analysis for biomonitoring activities, etc.). The project must be complete with the specification of all integrity constraints. The quality of the designed relational scheme must be verified, applying the principles of normalization theory. In particular, it will be necessary to establish whether the scheme is normalized with respect to the Boyce Codd normal form or with respect to the third normal form. If it is not normalized to the third normal form, propose a lossless, dependency-preserving decomposition for at least one non-normalized relation and indicate which normal form it satisfies.

II. Realization:

The purpose of this activity is the definition of the logical schema in PostgreSQL, its population and the implementation of some query operations on the database.

A. The definition of the logical scheme of the designed database must contain the specification of all the constraints identified in the design phase. For each constraint, the appropriate implementation method must be identified and indicated (CHECK or trigger type constraints). The implementation of only CHECK type constraints is required. As regards population, the database created must be populated with data sufficient to at least verify that the domain constraints expressed are verified and that in general the operations whose implementation is requested work correctly.

On this database it is required to create:

B. The definition of a view that provides some summary information for each biomonitoring activity: for each group and for the corresponding control group show the number of plants, the species, the garden in which the group is located and, on a monthly basis, the average value of the environmental and plant growth parameters (select at least three parameters, those considered most significant).

C. The following questions

to. determine the schools that, despite having funding for the project, have not included surveys in this school year;

b. determine the species used in all municipalities where there are schools participating in the project;

c. determine for each school the individual/class in the school who carried out the most surveys.