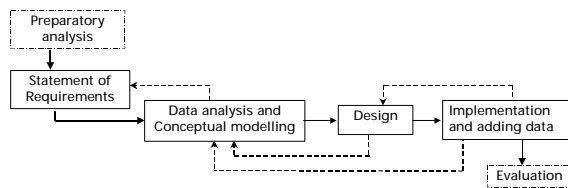# Bacteriocin Database Project

1. Project Overview and Approach

➤ **Solving a real problem, meeting a real need – with the aid of a relational database**

➤ In cooperation with Dr. Scannell from University College Dublin, the database subject matter was narrowed down to bacteriocin–producing microorganisms and related data (like its mode of action, uses in food, genetic determinants).

➤ Iterative database development process:

➤ Object Role Modelling was used in addition to Entity-Relationship modelling to uncover implicit assumptions and include them explicitly, leading to a comprehensive and flexible model for a communal database for bacteriocin researchers.

➤ Prototype database creation and new subject data induced further refinements of the model and design. User testing of the prototype confirmed validity of the model and design and provided suggestions for usability optimalisation.

➤ The final database installed at the customer site includes data extracted from 21 journal articles as a start, and easy to use forms, queries and reports created with InfoMaker features to access and, more importantly at this stage, to add more data.

➤ A *QuickStart Guide* was written to aid the customer and her students to be able to independently learn to use the database now and in the future.

2. Section of the ER Model

ER diagram

e.g. causesA **Disease**, or growsWith **StarterCulture**
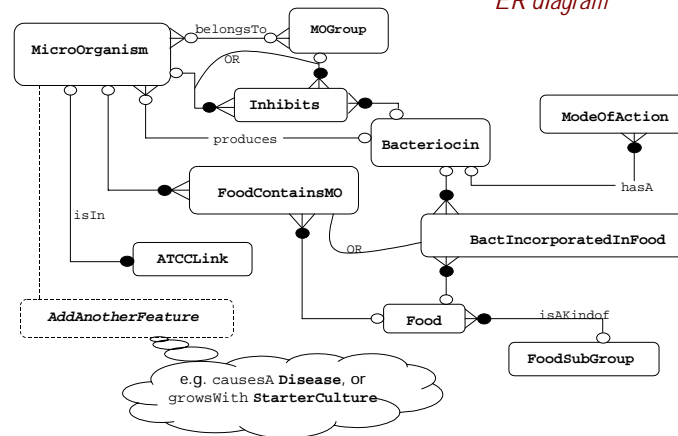
Entity types

```
MicroOrganism(IDMO,LatinNameFamily, LatinNameSub,
LatinNameSubSub)

Inhibits(IDMO, BactName, TargetProduces, MOGroupName,
Publications,)

Bacteriocin(BactName, BactGene, IsolationSource,
Publications, Comments)

ModeOfAction(ModeNr, BactName, PH, MoistureContent,
RestrictsFunctionality, OtherFactors, Status,
Publications, Comments)

FoodContainsMO(ProductName, IDMO, Publications,
Comments)

BactIncorporatedInFood(ProductName, BactName,
NatureOrAdd, Publications, Comments)

Food(ProductName, Comments)

FoodSubGroup(FoodSubGroup)

ATCCLink(IDMO, ATCCLinkWeb)

MOGroup(MOGroupName)
```

Note the centrality of **IDMO**, which identifies each microorganism. The **IDMO** can be used for any (new) relationship when expanding the database model, as indicated with **AddAnotherFeature**. Bacteriocins, produced by bacteria, have their own related features like the **ModeOfAction**, **BacteriocinType**, genes and **CommercialProduct**.
A **Literature** entity type is included to record each source of information, which can be searched on bacteriocins and/or its isolation source.
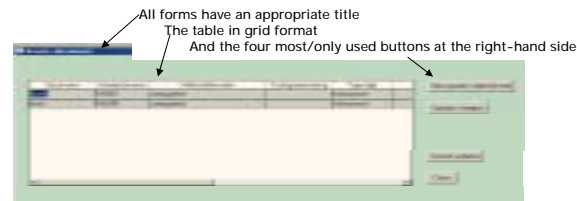
## 3. Design / Logical Model
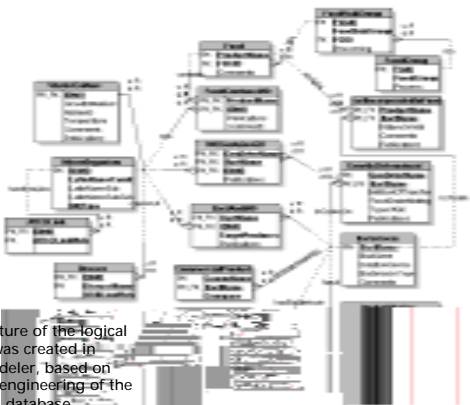


This picture of the logical model was created in VisioModeler, based on reverse engineering of the physical database.

Not all facets of the conceptual (ER) model could be represented in the design (most notably **MOGroup** with **Inhibits**). For efficiency, ease of use and/or lower importance, some entity types and relationships were consolidated (e.g. **BacteriocinType** as attribute in **Bacteriocin**, and **BactAndMO** combines inhibition of and production by microorganisms).

## 4. Implementation and Usability

Although the emphasis of the project was on database modelling, design and the general idea of a database software development process, the overall task does require "meeting a real need", which includes ensuring the customer can put the created database to its full advantage.

For each table, a form with a consistent interface was designed in InfoMaker and querying the database is possible via the ToolBox; all main user actions are described in a separate user guide.

All forms have an appropriate title
The table in grid format
And the four most/only used buttons at the right-hand side



The Query ToolBox



## 5. Outcome…

**Despite the known problems of biological data modelling, some of which encountered during this research, the customer's requirements were met. The bacteriocin database, with its integration of data from different life science and engineering fields, is the first of its kind. Therefore, the main value of this project is the contribution of these modelling aspects from an integrated food science perspective to bioinformatics research.**

### …and Future Research

➤ Adding contents, based on food science, microbiology and genetics journal articles;

➤ Expanding the amount of entity types (other microbiological produce) and attributes;

➤ Separation of literature references into a connected database and/or possible database connectivity with primary biological databases;

➤ To represent gradations in inhibition of microorganisms as opposed to a yes/no relationship with bacteriocins and, together with the mode of action of bacteriocins, the interdependencies of the relevant environmental factors.

However, viewed in a wider perspective, the database-focussed bioinformatics community will need to come to an agreement with life science researchers on figuring out computing-suitable structures, nomenclatures and conceptual modelling approaches based on preferably one type of ontology, or at least several that can communicate with one another without duplication, inconsistencies and 'translational' problems between as of yet largely incompatible data(base) models, designs and implementations.