

# How to Survive Real-World Projects as a Mathematician

Lessons and experiences from 10 year of industry projects



ORSUM/MASCOS Seminar  
Melbourne, October 2010

**DFG Research Center Matheon**  
Mathematics for key technologies

**Thorsten Koch**  
Zuse Institute Berlin

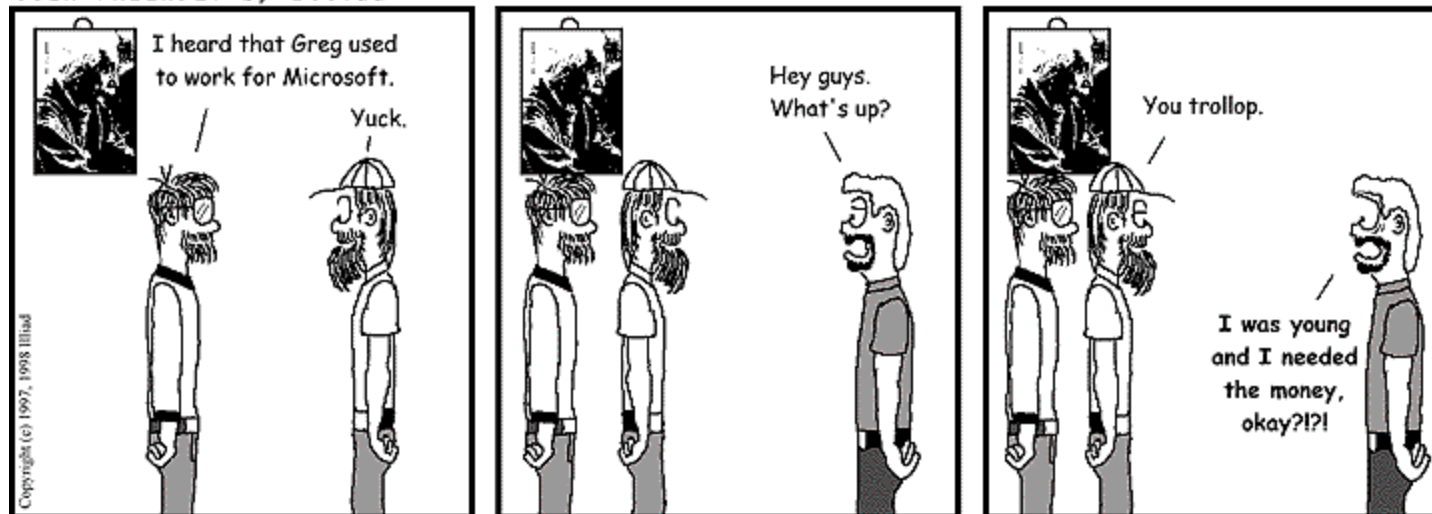




# Why do real-world projects at all?

- ▶ take a lot of time and effort
- ▶ can fail miserably
- ▶ often lack theoretical appeal
- ▶ results may be hard to publish
- ▶ have impact in the real-world
- ▶ challenging because rules are set
- ▶ improve something people use
- ▶ somebody actually cares about the result

USER FRIENDLY by Illiad



**WRONG ANSWER!**



How to convince the industry people that you can help them:

- They are the specialists for the topic not you.  
You are not even an engineer.
- Be aware they do not want a result that says:  
We can prove there exists a unique solution.
- If asked, how much you can improve on the current solution,  
the correct answer is 15%  
(see G. Dueck, DMV-Mitteilungen, 2003, 44-45)
- Even if you know something about their business, regardless  
whom you ask, they will tell you : "We are special"  
Corollary: Since everybody is special, they are all equal.
- If you try to convince them by showing something similar, they  
might have a very focused view on their particular problem.



Sometimes a company will suggest to do a pilot project first:

- The unspoken expectation is that you put in more resources than what you are paid for.
- Chances for a continuation project are as good with or without a pilot project.
- You will have trouble to get up your prices again afterwards.
- You can do this, but the default has to be the continuation. Just suggest the right to drop out at a certain point in case of failure or unsatisfying results.



- Intellectual property rights (patents)
- 3<sup>rd</sup> party code
- Don't do maintenance
- Right to publish
- Right to give talks
- Right to cooperate with others
- Right to continue afterwards with others (competitors)

**Remember: The contract is basically useless, as you will never sue and can do little later on.**



Removed due to legal reasons.



Removed due to administrative reasons.



**DATA**





# DATA



If you start thinking about modelling the problem, you should immediately check if the necessary data is potentially available.

If you ask, there are two possible answers:

- a. We do not have the data
- b. We have the data

**Usually, both are somewhat wrong.**



Combinatorial Optimization at Work II took place at ZIB from September 21 to October 9, 2009 with 105 participants from 23 countries.

We wanted to compute the seat allocation for the lecture hall.

To do this we required every participant to state their preferences.

Everyone should send an email with a data file.

Lets see how long it took...



ASCII text with only a LF (ASCII 10) as line separator.

Fields are separated by a single space (ASCII 32)

Line 1: **ParticipantNo** **HasLaptop** **EmailAddress**

e.g.            **67 1 koch@zib.de**

0 = has no Laptop, 1 = has a Laptop

Lines 2-???: **SeatNumber** **PreferenceValue**

- Seat numbers start down at the low entrance, left to right, row by row.
- The highest numbered seat is at the window side at the top.
- Count only seats that are physically there.
- The seat numbers in the file should be monotonically increasing.
- The preference values should be between 0 and 100.

e.g.            **12 55**  
                  **13 40**  
                  **14 35 ...**



Allowed values are between 0 and 100

Only seats which are not available for the participants are allowed to get a value of 0

All numbers 1-100 have to be used at least once

The average has to be between 40-60

The difference to an adjacent seat has to be  $< 40$

The difference to a neighboring seat has to be  $< 20$

The data should not be randomly generated



Lines ???-???: **ParticipantNo PreferenceOffset**

List indicating persons which you would like or not like to be your seat neighbor. (You have to know the ParticipantNo of the person.)

- A ParticipantNo of 0 indicates an empty seat.
- The PreferenceOffset is between -20 and 20 and will be added to your PreferenceValue if the person with the given ParticipantNo is your neighbor.

e.g.           55 17  
              27 -5  
              72 8  
              0 -10 ...

- This list can have as many entries as you like, but there should be at least 2 entries, and the occurring participant numbers have to be unique and valid.



Submission of this file is required for the course

The name of the file has to be *ParticipantNo.txt*

It should be **attached** to an email

Send the email to [koch@zib.de](mailto:koch@zib.de)

The subject of the email should be

CO@Work: SeatData for *ParticipantNo*

*Please, as soon as possible.*



|                     |                |
|---------------------|----------------|
| Mails received      | : 13           |
| Different Subjects  | : 4 (10 1 1 1) |
| Wrong field spacing | : 4            |
| Seat counts         | : 2 (12 1)     |
| Missing data        | : 1            |
| Too much data       | : 1            |
| Ok, from first view | : 5 out of 13  |





|                     |                    |
|---------------------|--------------------|
| Mails received      | : 23               |
| Different Subjects  | : 6 (17 2 1 1 1 1) |
| Wrong field spacing | : 4                |
| Seat counts         | : 4 (19 1 1)       |
| Missing data        | : 2                |
| Too much data       | : 0                |
| Ok, from first view | : 10               |
| Corrected           | : 1                |

## Add to the specification:

A seat without a desk is not allowed for the participants  
Seats with a 0 preference value are not relevant for the adjacency/neighborhood difference rules.



|                     |      |
|---------------------|------|
| Mails received      | : 37 |
| Wrong subject       | : 11 |
| Wrong field spacing | : 8  |
| Strange seat counts | : 5  |
| Missing data        | : 2  |
| Corrected           | : 3  |



|                        |                      |
|------------------------|----------------------|
| Mails received         | : 47                 |
| Data sets              | : 41 (6 corrections) |
| Wrong subject          | : 12                 |
| Wrong attachment name: | 2                    |
| Wrong line separator   | : 29                 |
| Wrong field separator  | : 10                 |
| Pref. value not used   | : 11                 |
| Other Errors           | : 1                  |
| Number of seats        | : 153 - 181          |
| No complains so far    | : 4                  |



|                        |             |
|------------------------|-------------|
| Mails received         | : 79        |
| Data sets              | : 64        |
| Wrong subject          | : 16        |
| Wrong attachment name: | 2           |
| Wrong line separator   | : 45        |
| Wrong field separator  | : 11        |
| Pref.value not used    | : 22        |
| Other Errors           | : 2         |
| Number of seats        | : 153 - 181 |
| No complains so far    | : 8         |



|                        |       |
|------------------------|-------|
| Mails received         | : 104 |
| Data sets              | : 76  |
| Wrong subject          | : 18  |
| Wrong attachment name: | 2     |
| Pref. value not used   | : 19  |
| Neighbor difference    | : 21  |
| Wrong no/seq. seats:   | : 10  |
| Wrong 0 seats          | : 20  |
| No complains so far    | : 10  |



# Overview of Errors in Data

|    | E7 | E10 | E11 | E12 | E13 | E14 | E16 |
|----|----|-----|-----|-----|-----|-----|-----|
| 5  |    |     |     |     |     |     | X   |
| 6  |    |     |     |     |     |     | X   |
| 12 |    |     |     |     | X   |     | X   |
| 13 |    |     |     |     |     |     | X   |
| 16 |    |     |     |     |     |     | X   |
| 18 |    |     |     |     | X   |     | X   |
| 19 |    |     |     |     |     | X   | X   |
| 20 |    |     |     |     |     | X   |     |
| 23 |    |     |     |     | X   |     |     |
| 24 |    |     |     |     |     | X   |     |
| 26 |    |     |     |     |     |     | X   |
| 27 |    |     |     |     |     | X   |     |
| 36 |    |     |     |     |     | X   |     |
| 42 |    |     |     |     | X   |     |     |
| 45 |    |     | X   | X   | X   | X   | X   |
| 47 |    |     |     |     | X   |     |     |
| 53 |    |     |     |     | X   |     |     |
| 59 |    |     |     |     |     | X   |     |
| 63 |    |     | X   |     | X   | X   |     |
| 64 |    |     | X   |     | X   | X   | X   |
| 71 |    |     |     |     | X   | X   |     |

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 prefval missing

E14 neighbour diff

E16 seat not 0

|     | E7 | E10 | E11 | E12 | E13 | E14 | E16 |
|-----|----|-----|-----|-----|-----|-----|-----|
| 77  |    |     |     |     |     |     | X   |
| 78  | X  |     | X   |     |     | X   | X   |
| 81  |    |     |     | X   | X   |     | X   |
| 98  |    |     |     |     | X   |     |     |
| 99  | X  |     | X   |     |     | X   |     |
| 103 |    |     |     |     | X   | X   |     |
| 107 |    |     | X   |     |     | X   | X   |
| 108 |    |     | X   |     |     | X   | X   |
| 111 |    |     |     |     |     |     | X   |
| 121 |    |     |     |     |     |     | X   |
| 128 |    |     | X   |     |     | X   | X   |
| 129 |    | X   |     |     |     |     |     |
| 134 |    |     | X   | X   | X   | X   |     |
| 135 |    |     |     |     | X   |     |     |
| 137 |    | X   |     |     | X   | X   | X   |
| 139 |    |     |     |     | X   |     | X   |
| 145 | X  |     | X   |     | X   | X   |     |
| 160 |    |     |     |     |     | X   |     |
| 166 |    |     |     |     | X   | X   |     |

**Please correct and resubmit**



|                        |       |
|------------------------|-------|
| Mails received         | : 144 |
| Wrong subject          | : ~23 |
| Wrong attachment name: | 4     |
| <br>                   |       |
| Data sets              | : 92  |
| To be corrected        | : 28  |
| Missing                | : 6   |
| <br>                   |       |
| Pref. value not used   | : 14  |
| Neighbor difference    | : 18  |
| Wrong no/seq. seats    | : 2   |



|     | E7 | E10 | E11 | E12 | E13 | E14 |
|-----|----|-----|-----|-----|-----|-----|
| 12  |    |     |     |     | X   |     |
| 18  |    |     |     |     | X   |     |
| 23  | X  |     | X   |     |     | X   |
| 24  |    |     |     |     |     | X   |
| 27  |    |     |     |     |     | X   |
| 45  |    |     |     |     | X   | X   |
| 47  |    |     |     |     | X   |     |
| 63  |    |     | X   |     | X   | X   |
| 71  |    |     |     |     | X   | X   |
| 78  | X  |     | X   |     |     | X   |
| 79  |    | X   | X   | X   | X   |     |
| 103 |    |     |     |     |     | X   |
| 107 |    |     | X   |     |     | X   |
| 108 |    |     | X   |     |     | X   |
| 110 |    | X   |     |     |     |     |
| 114 |    |     |     |     |     | X   |
| 118 |    |     |     |     | X   | X   |
| 128 |    |     | X   |     |     | X   |
| 134 |    |     | X   | X   | X   | X   |
| 135 |    |     |     |     | X   |     |
| 136 |    |     |     |     |     | X   |
| 137 |    | X   |     |     | X   | X   |
| 138 |    |     |     |     | X   |     |
| 139 |    |     |     |     | X   |     |
| 160 |    |     |     |     |     | X   |
| 166 |    |     |     |     | X   | X   |

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 preaval missing

E14 neighbour diff

**Please correct and resubmit**





|                           |       |
|---------------------------|-------|
| Mails received            | : 159 |
| Wrong subject             | : ~26 |
| Wrong attachment name     | : 4   |
| <br>                      |       |
| Data sets                 | : 94  |
| To be corrected           | : 18  |
| Missing                   | : 4   |
| <br>                      |       |
| Preference value not used | : 9   |
| Neighbor difference       | : 14  |
| Wrong no/sequence seats   | : 3   |



|     | E7 | E10 | E11 | E12 | E13 | E14 |
|-----|----|-----|-----|-----|-----|-----|
| 18  |    |     |     |     | X   |     |
| 24  |    |     |     |     |     | X   |
| 27  |    |     |     |     |     | X   |
| 45  |    |     |     |     | X   | X   |
| 63  |    |     |     |     | X   |     |
| 71  |    |     |     |     | X   | X   |
| 78  | X  |     | X   |     |     | X   |
| 79  |    | X   | X   | X   | X   |     |
| 103 |    |     |     |     |     | X   |
| 107 |    |     | X   |     |     | X   |
| 108 |    |     | X   |     |     | X   |
| 114 |    |     |     |     |     | X   |
| 118 |    |     |     |     | X   | X   |
| 128 |    |     | X   |     |     | X   |
| 134 |    |     | X   | X   | X   | X   |
| 136 |    |     |     |     |     | X   |
| 137 |    | X   |     |     | X   | X   |
| 138 |    |     |     |     | X   |     |

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 prefval missing

E14 neighbour diff

**Please correct and resubmit**



|                           |       |
|---------------------------|-------|
| Mails received            | : 166 |
| Wrong subject             | : ~28 |
| Wrong attachment name     | : 4   |
| <br>                      |       |
| Data sets                 | : 95  |
| To be corrected           | : 18  |
| Missing                   | : 3   |
| <br>                      |       |
| Preference value not used | : 7   |
| Neighbor difference       | : 14  |
| Wrong no/sequence seats   | : 3   |



|     | E7 | E10 | E11 | E12 | E13 | E14 |
|-----|----|-----|-----|-----|-----|-----|
| 24  |    |     |     |     |     | X   |
| 27  |    |     |     |     |     | X   |
| 45  |    |     |     |     | X   | X   |
| 71  |    |     |     |     | X   | X   |
| 78  | X  |     | X   |     |     | X   |
| 79  |    | X   | X   | X   | X   |     |
| 92  |    |     |     |     | X   | X   |
| 107 |    |     | X   |     |     | X   |
| 108 |    |     | X   |     |     | X   |
| 114 |    |     |     |     |     | X   |
| 118 |    |     |     |     | X   | X   |
| 128 |    |     | X   |     |     | X   |
| 134 |    |     | X   | X   | X   | X   |
| 136 |    |     |     |     |     | X   |
| 137 |    | X   |     |     | X   | X   |

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 prefval missing

E14 neighbour diff

**Please correct and resubmit**



|                           |       |
|---------------------------|-------|
| Mails received            | : 172 |
| Wrong subject             | : ~31 |
| Wrong attachment name     | : 4   |
| <br>                      |       |
| Data sets                 | : 95  |
| To be corrected           | : 13  |
| <br>                      |       |
| Preference value not used | : 5   |
| Neighbor difference       | : 13  |
| Wrong no/sequence seats   | : 2   |



The subject of the email should be  
**CO@Work: SeatData for *ParticipantNo***

CO@Work: SeatData for 022  
CO@Work:SeatData for 222  
CO@Work:SeatDatafor222  
CO@work: SeatData for 222  
CO@Work: Seat Data for 222  
Co@Work: SeatData for 222  
CO@Work: SeatData for Participant222  
CO@Work: SeatData for ParticipantNo  
Co@Work: SeatData for Participan222  
CO@WORK: seatdata for 222  
COatWork: SeatData for 222  
COatWork for 222  
SeatData for 222  
SeatData for ParticipantNo 222  
set data for participant number 222  
data set participant number 222  
Sitting assignment  
Seats assignment



|     | E7 | E10 | E11 | E12 | E13 | E14 |
|-----|----|-----|-----|-----|-----|-----|
| 24  |    |     |     |     |     | X   |
| 27  |    |     |     |     |     | X   |
| 45  |    |     |     |     | X   | X   |
| 71  |    |     |     |     | X   | X   |
| 78  | X  |     | X   |     |     | X   |
| 92  |    |     |     |     | X   | X   |
| 107 |    |     | X   |     |     | X   |
| 108 |    |     | X   |     |     | X   |
| 114 |    |     |     |     |     | X   |
| 128 |    |     | X   |     |     | X   |
| 134 |    |     | X   | X   | X   | X   |
| 136 |    |     |     |     |     | X   |
| 137 |    | X   |     |     | X   | X   |

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 prefval missing

E14 neighbour diff

**Sorry,  
too late to correct!**

Wrong line 1: 81, 129



Removed due to confidentiality.

- If you do not find at least 1 error per 100 data sets you are not looking hard enough.
- Usually the data changes all the time.
- They might not want to give it to you.





## The data might just not exist.

We all know that centrally planned economies did not do too well.

One of the (major) reasons was that the assumptions, i.e. the data they used to make their plans were faulty.

Now the big companies just try the same 😊

The first result of an optimization project is usually to improve the quality of planning data available at the company.



Find out what the real objective is. Usually it is not what you have been told in the beginning.

Sometimes it is difficult because the cost impact when optimizing virtual or already existing structures is unclear.

Often you have to compare apples and pies.

Non linear objectives would help for soft constraints, but ...



More often than not real-world problems are rather big.

At least compared to “academic toy examples”

In all projects I have done so far, the key to success was proper preprocessing of the data. Since it is real-world, many things will be obvious to decide. Throw them away before they make trouble.



The traveling salesman problem is to mathematical programming what chess is to artificial intelligence: thoroughly useless and fiercely competitive sport that serves as a testing ground of your techniques. --- Vasek Chvatal

Do you know what

**Zone defining Origins** (verzonende Ursprünge) are?

We have one location where there is still the old XXX which behaves totally different.

Problem: Decide what you have to model and what to ignore (or fix later)



# Commercial Break



## The 8th International Conference on Integration of Artificial Intelligence and Operations Research Techniques in Constraint Programming for Combinatorial Optimization Problems

23-27 May 2011 at ZIB

<http://cpaior2011.zib.de>





## Mathematical Programming Computation

MPC will publish original research articles covering computational issues in mathematical programming.

Articles report on innovative software, comparative tests, modeling environments, libraries of data, and/or applications.

A main feature of the journal is the inclusion of accompanying software and data with submitted manuscripts. The journal's review process includes the evaluation and testing of the accompanying software. Where possible, the review will aim for verification of reported computational results.

Visit: <http://mpc.zib.de> and <http://springer.com>

Editor-in-Chief: William Cook

General Editor: Thorsten Koch

Area Editors: D. Bienstock, R. Fourer, A. Goldberg, S. Leyffer, J. Linderoth,  
A. Martin, G. Reinelt, K. Toh



|                    |                  |                   |
|--------------------|------------------|-------------------|
| Tobias Achterberg  | Jordi Castro     | Leonardo B. Lopes |
| David Applegate    | Daniel Espinoza  | Todd S. Munson    |
| Erling D. Anderson | Armin Fügenschuh | Dominique Orban   |
| Oliver Bastert     | Andreas Grothey  | Ted Ralphs        |
| Pietro Belotti     | Zonghao Gu       | Mohit Tawarnalani |
| Hande Y. Benson    | William Hart     | Stefan Vigerske   |
| Andreas Bley       | Keld Helsgaun    | Richard A. Waltz  |
| Brian Borchers     | Benjamin Hiller  |                   |





Optimality is usually not required.

Extremal Solutions vs. 80%

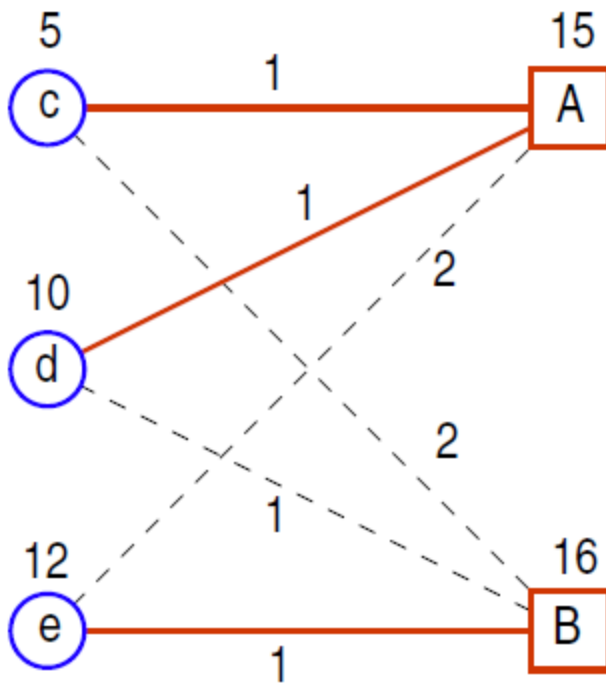
**It is very important that the solution is 2-optimal.**

To compute a gap or prove infeasibility is important.

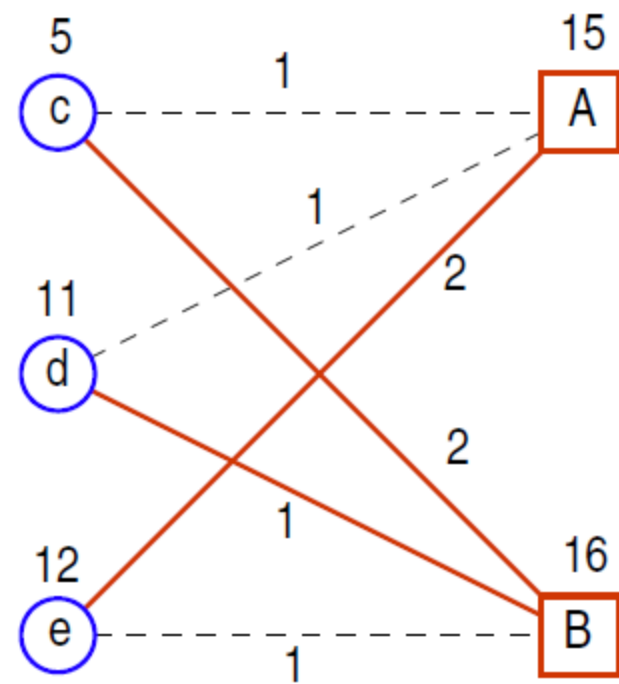
What does it mean? You are proveably global optimal but the former solution the company used is better.

Strange solutions may have several reasons:

- ▶ Anomalies (errors) in the data
- ▶ Differences between model and reality
- ▶ Reaching of capacity or cost thresholds
- ▶ The result is just different than expected



(a) Optimal solution



(b) After a small change



- ▶ The bigger the company, the more unstable is the department.
- ▶ If the project takes too long they may lose interest.
- ▶ This is research not development.
- ▶ Milestones in the beginning are hard to meet for research projects.
- ▶ Check the solutions, we do errors, too.



I will survive



**Thank you very much!**



# Questions?