Software Development in Agricultural Computer Science: Status and Perspectives

Eildert Groeneveld

eg@tzv.fal.de

Rodica, April 15th. 2004

Institute for Animal Breeding, FAL, Mariensee
Outline

1. Objectives in Computer Science in Agriculture
2. The Framework
3. Thesis
4. the “so what?” projects
5. examples of dead-end developments
6. possible solutions and development paths
7. examples of the OS development model
8. Open Source development
9. outlook
Objectives in Agr. Computer Science

Facilitate IT solutions in Agriculture

- quick development and implementation of new solutions
- requires research and development

Essential framework:
- cumulative mode of development
- leads to accelerated development
- is based on original research

Progress resulting from the general development in SW/HW does not count
The Framework

Information processing in Agriculture:

- in the scientific area:
  - design and evaluation of experiments
  - development of computing algorithms
  - well defined inputs -> problem solution -> Output (IPO)
The Framework ff

Information processing in Agriculture:

- on farm area:
  - solutions to local problems: optimization of feed rations (IPO)
  - area of global information: collection, management and evaluation
  - merging data from various sources:
    - lots of land, soil samples
    - performance recording: milk, growth, reproduction
- across farms:
  - aggregation of data from many sources (InfoSys)
  - in animal production: BLUP genetic evaluation
The Framework

Information processing in Agriculture:

- animal breeding, crop production, agricultural engineering, agricultural economics
- different kind of systems (IPO \& InfoSys):

  IPO: production of data -> problem solution -> decision -> throw data away
  InfoSys: integrated data utilization: breeding, milk recording, labs, AI, utilization of the WEB, networking of farms
IPO vs InfoSys

conceptual differences:

IPO:

- simple well defined inputs
- complex processing but usually in a simple program environment (F90, C, ...)
- well defined outputs
- processing layer is disconnected from the inputs
Examples IPO

- statistical procedures through use of packages (SAS, R...)
- analysis of inbreeding (SW)
- computation of BLUP (SW)
- estimation of covariance components (SW)
- etc.
IPO vs InfoSys ff

conceptual differences:

InfoSys:
- many sources of data from different origin/systems
- as a result: different SW
- as a result: different operating systems
- includes data transfer
- different data structures among different problems (organizations)
- everything is connected with everything
Examples InfoSys

- gene banks in plants and animals
- e-Commerce Systems
- query and information systems
- integrated data bases: plants/animals
- many components:
  - Webserver,
  - DBMS,
  - programming languages,
  - OS
Thesis

IPO:
1. many developments with large impact: DFREML, MTDFREML, ASREML, VCE, ABTK, PEST

InfoSys:
1. many developments that solve ONE problem
2. are thus irrelevant for the industry as a whole:
3. the “So What?” projects.
The Problem: An Example

WCGALP Australia in Section:
"Computing Techniques in Animal Breeding"

- total of 26 contributions
- 17 IPO
- 4 InfoSys:
  - NZ sheep improvement scheme
  - QTL detection
  - WEB Animal Recording Scheme
  - AMNOS Dairy Sheep Management
  - all SW

Software Development in Agricultural Computer Science: Status and Perspectives – p. 12
Analysis of the Problem: own examples

- 1970: central database for the BHZP:
  - HP-1000
  - network type database
  - electronic on-farm data recording
  - electronic automated data transfer
  - remote program execution

- Effect:
  - solves the BHZP problem, but no other despite publication
  - sorry to say: SW project
Analysis of the Problem: own examples ff

- End 1980:
  - new attempt at generalization:
  - ORACLE
  - SUN-OS
  - pig breeding program Slovenia
  - transferred to Chinese Swine Herd at U of I (USA):
  - no transfer beyond
  - solved a problem in Slovenia and U of I but nothing else despite Master’s thesis
  - sorry to say: SW project
A Key Experience

- meeting of developers of integrated databases
- different countries
- different external circumstances
- remarkable:
  - different systems and tools (dBase, Clipper, Oracle...)
  - totally different development paths
  - every system a completely new development
    without reference to previous work (literature?)
  - no way to use software across
  - nothing worked really well
Critical Question

How many development projects in the different areas are of the kind: “So What”? 
What to do?

- the basic problem:
  - solution of ONE problem
  - thus: not transferable
  - because solution of one problems includes all side conditions:
    - operating system
    - RDBMS
    - system tools and software
  - this is hardly ever an issue in ONE problem
  - but strong restriction on transfer
  - software is generally not available
What to do? ff

- solution can only lie in:
  - generalization: find the Poodle’s core
  - also: never develop for ONE case, but always for many
  - parallel implementation
  - implement a layered system which allows replacing components (non monolithic)
  - release software
A New Development Path

- started as “Integrated Database of Performance Records in Pig Populations”
- any animal identification system
- any data structure
- installation free of licencing costs
- scales from on-farm to national
- development of procedures for development and migration
- development of software toolbox
- development of general evaluation software
A New Development Path ff

- avoid proprietary software
- use standards (SQL-99)
- use Open source
- developer group
- implementation in at least two environments
Development

- **Strategy/Procedure**
  - can be used independently from software
  - ⇒ migration steps of an information system
- **Software**
  - for the implementation of the procedures
- generic solution for BR
Closed or Open Development?

- IPO can work without program modification
  - InfoSys: different data structures require software modification
    - System without source code not conceivable
- closed development:
  - few people have the knowledge
  - restricted pool for future development
- need to create development capacity
  - OpenSource
Open Source Development

The Cathedral & the Bazaar
(ERIC S. RAYMOND)

“given enough eyeballs, all bugs are shallow”

big project:

- EMACS, LINUX, FETCHMAIL
- KDE, GNOME

requires infrastructure:

- Internet connection
- Software management system (CVS)
Open Source: Advantages

- get help with debugging
- expand developers base
- enhances future system safety through expanded knowledge base
- help during conceptual development
- cost effective developer and test capacity via the Internet
- the world is the limit (cooperation space)
Implementers

- Institute of Animal Production, Irene, South Africa
- Landesanstalt für Tierzucht, Köllitsch
- University of Ljubljana, Slovenia
- Research Institute Nitra, Slovakia
- University Debrecen, Hungary
- Veterinary Academy Kaunas, Lithuania
- University Stara Zagora, Bulgaria
- FAO
- University Göttingen
- TZ, Mariensee
Current Projects

- PISSA South Africa, Pigs, Cheetah
- Beef South Africa, Cattle
- PIS Lithuania, Pigs
- CryoDB Netherlands, Germany, Genebank
- MINIPIGS Göttingen, pigs
- MOLABIS Germany, DNA Material
- SCHAF-OT Germany, Sheep
- Saxonian Pig System, Germany
- EFABIS (FAO, F, PL, D), Biodiversity Database
- a number in Slovenia: Rabbits, Horses, Cattle, Pigs, ..
Platforms create Utility

- many developers in CC, Research Inst. but SW generally not usable in InfoSys
- on generic platforms software can be used:
  - Population report
  - ZWISSS
  - ...
- cumulative development model
Population Report 1

PISSA
Population Structure Report

Frits Voordewind
Institute for Animal Improvement
ARC Pbag X2 Irene 0062
13, Januar 2003

The PISSA
database
1 General

2 Number of sires and dams in reproduction by year

This report lists the number of boars and sows used in reproduction. We have one line per year (see Table 1). The first part of the table lists the sires followed by sows. Every section has three columns. They are parents (boars and sows):

- 1. used in service: here we count the number of sows and boars who participated in services during the corresponding year.
- 2. in litters: this is the total number of sows and boars with litters in the given year.
- 3. gives the number of boars and sows having selected offspring in the given year. Selection is defined as having a first service. Thus, this is the number of sows and boars with a first service in the given year.

Table 1: Number of sires and dams in reproduction by year (CW)

<table>
<thead>
<tr>
<th>Year</th>
<th>number of sires</th>
<th>services</th>
<th>litters</th>
<th>offspring, sel.</th>
<th>number of dams</th>
<th>services</th>
<th>litters</th>
<th>offspring, sel.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1982</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1983</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1984</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1985</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1986</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>10</td>
<td>11</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>10</td>
<td>5</td>
<td>6</td>
<td>24</td>
<td>20</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>11</td>
<td>12</td>
<td>8</td>
<td>27</td>
<td>24</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>14</td>
<td>10</td>
<td>11</td>
<td>29</td>
<td>26</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>22</td>
<td>19</td>
<td>13</td>
<td>48</td>
<td>46</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>26</td>
<td>19</td>
<td>18</td>
<td>58</td>
<td>48</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>31</td>
<td>33</td>
<td>16</td>
<td>71</td>
<td>64</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>21</td>
<td>23</td>
<td>23</td>
<td>78</td>
<td>76</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>28</td>
<td>28</td>
<td>15</td>
<td>77</td>
<td>83</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>26</td>
<td>25</td>
<td>13</td>
<td>50</td>
<td>53</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>10</td>
<td>14</td>
<td>9</td>
<td>33</td>
<td>42</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td></td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Population Report 2

MINIPIGS
Population Structure Report

Eildert Groeneveld
Institute for Animal Science and Animal Husbandry
31535 Neustadt Mariensee
7. Juli 2003

MINIPIGS Göttingen
Table 1: Number of sires and dams in reproduction by year (go)

<table>
<thead>
<tr>
<th>Year</th>
<th>number of sires</th>
<th>number of dams</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>services</td>
<td>litters</td>
</tr>
<tr>
<td>1984</td>
<td>69</td>
<td>38</td>
</tr>
<tr>
<td>1985</td>
<td>83</td>
<td>62</td>
</tr>
<tr>
<td>1986</td>
<td>71</td>
<td>59</td>
</tr>
<tr>
<td>1987</td>
<td>71</td>
<td>59</td>
</tr>
<tr>
<td>1988</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>1989</td>
<td>62</td>
<td>53</td>
</tr>
<tr>
<td>1990</td>
<td>49</td>
<td>39</td>
</tr>
<tr>
<td>1991</td>
<td>62</td>
<td>58</td>
</tr>
<tr>
<td>1992</td>
<td>62</td>
<td>60</td>
</tr>
<tr>
<td>1993</td>
<td>47</td>
<td>46</td>
</tr>
<tr>
<td>1994</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>1995</td>
<td>41</td>
<td>30</td>
</tr>
<tr>
<td>1996</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>1997</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>1998</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>1999</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>2000</td>
<td>24</td>
<td>14</td>
</tr>
<tr>
<td>2001</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>2002</td>
<td>12</td>
<td>23</td>
</tr>
</tbody>
</table>
is SW development in InfoSys or IPO research?
I think it is. Need to change attitudes.
general problem with OS projects
may be: we are part of the problem?
Summing Up

Commercial Development: solve ONE problem

Research in CS: develop generic solutions

- research in this area is not uncontested
- clean definition of layers
- Open Source
- avoid proprietary solutions
- push for a cumulative development model