

SYRCoDIS: A Path to Quality and Success

Boris Novikov

Prof., Sub-Dept. Analytical information systems

Saint Petersburg University, Russia

b.novikov@spbu.ru

Russian (DBIS) Research is not Competitive. Why?

Very often results are not competitive

Even good results are not presented properly

Anonymous reviewing is not in use

The word 'peer' cannot be translated into Russian

The position of unrecognized genius is very comfortable

It is easy to blame political or economical situation

SYRCoDIS: Motivation and Objectives

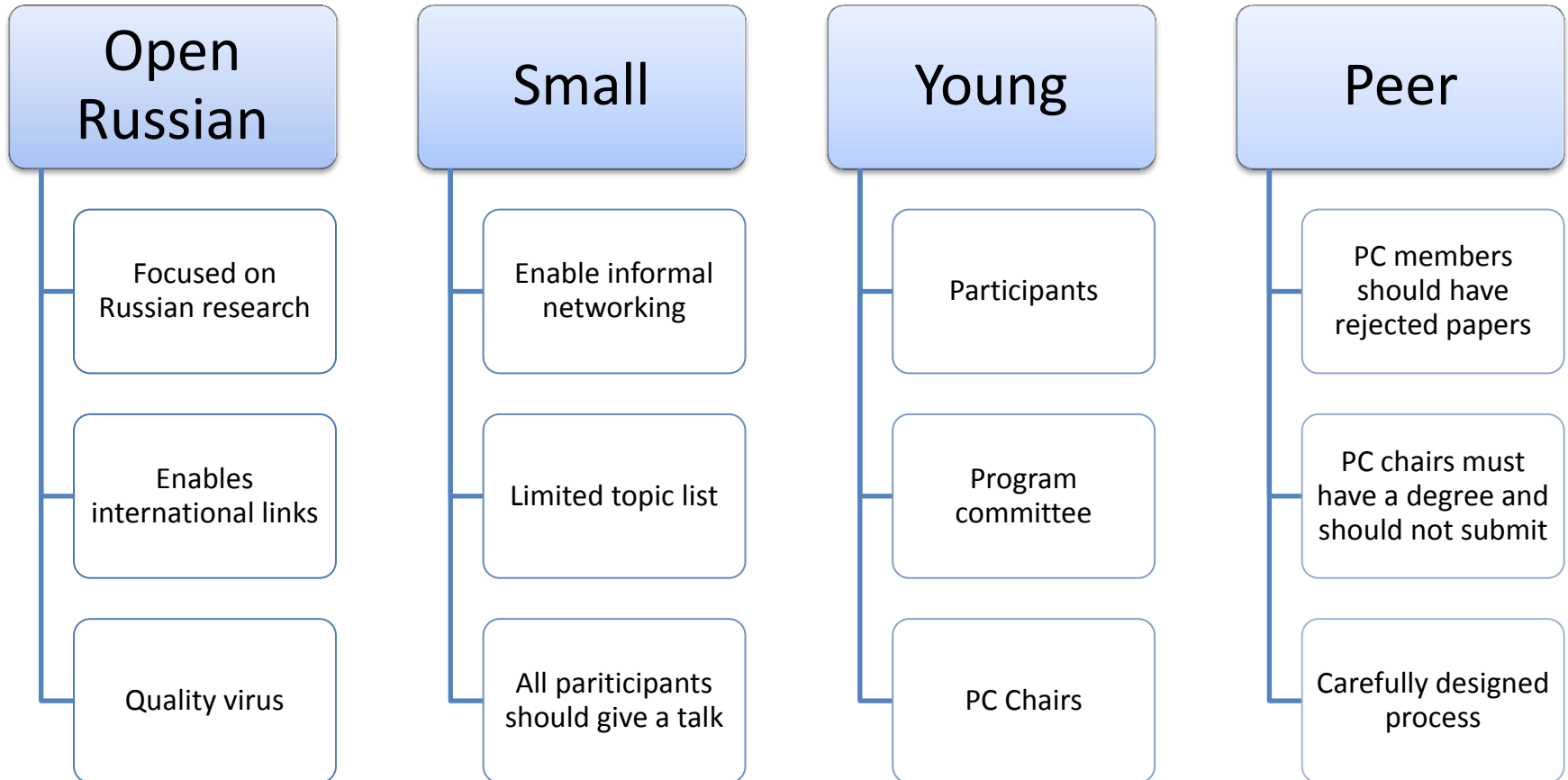
Comparing Intl and Russian conferences

- Different reviewing/selection process
- Avg. age: VLDB, SIGMOD etc.: 30 – C35, RCDL: 65+

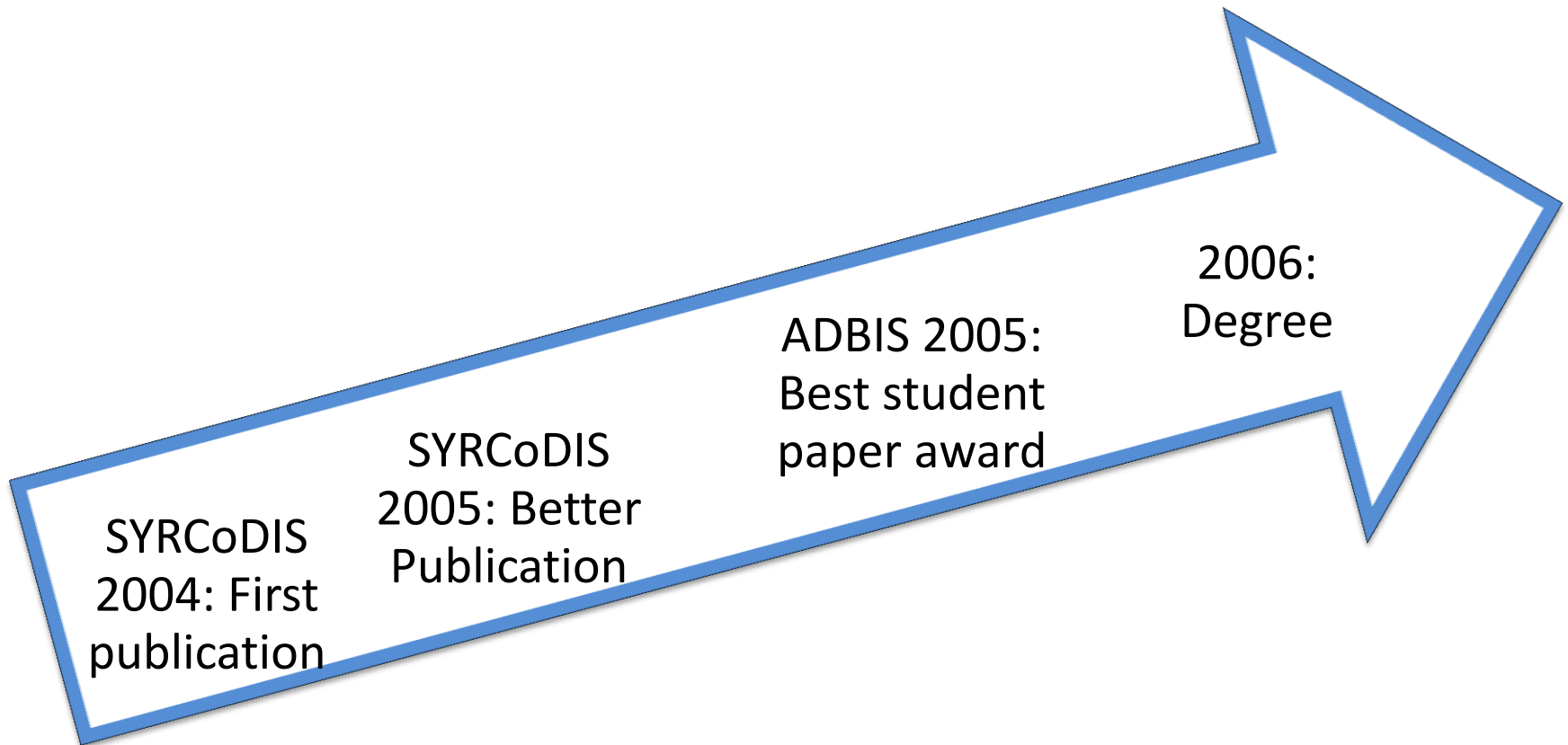
Objectives:

- A bridge to quality and competitiveness
- Implant peer reviewing tradition and skills
- Networking with external research community

SYRCoDIS: The Design



A Success Story

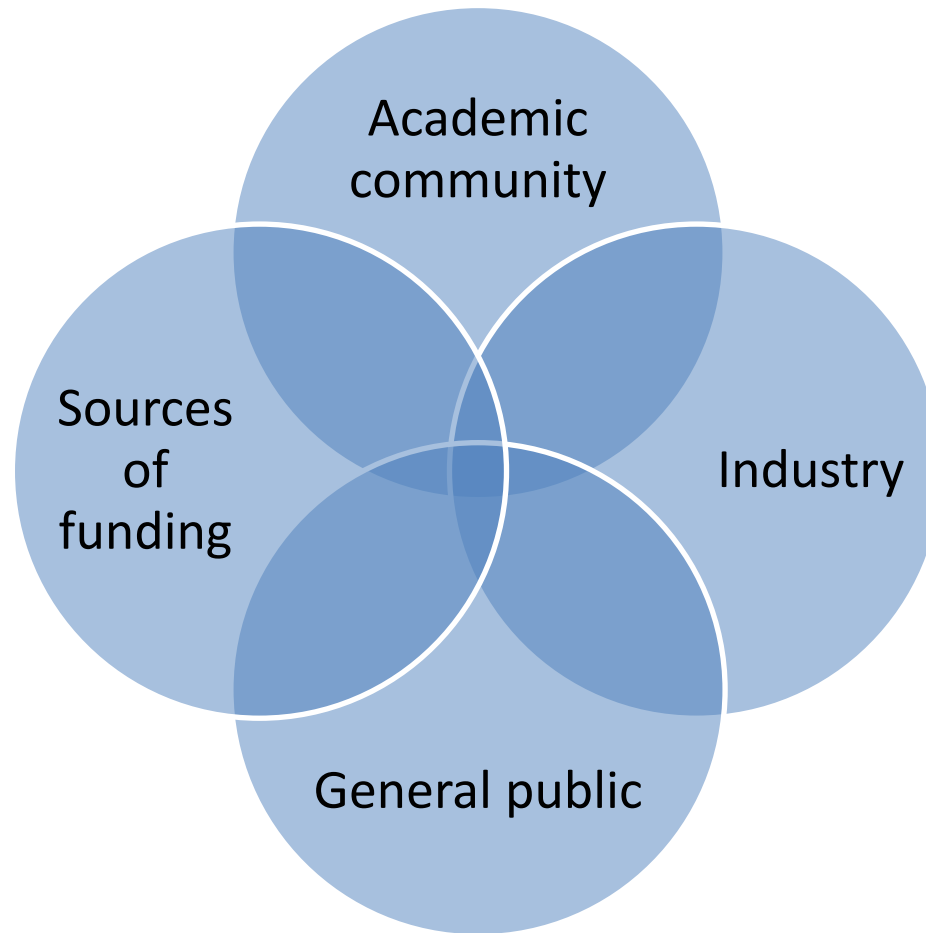


Can High Quality Research be
Really Useful?

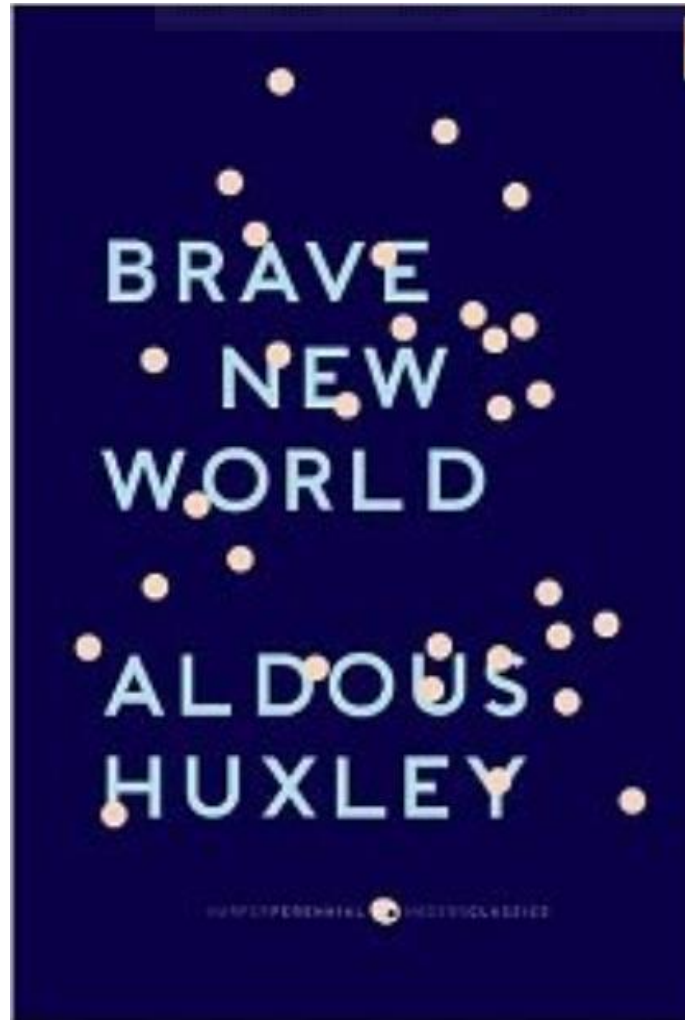
An Immediate answer: Certainly NO.

	Research	Production
The outcome	Knowledge	Product
Novelty	Is a must	Not necessary
Timeliness	Desirable	Is a must
Balance between features/parts	Not needed	Is a must
Reliability	Not needed	Expected
Relative cost	x	Approx. 10x

Viewpoints



5000 Repetitions



What is a Success?

- Academic research
 - Publications
 - degrees
- Industrial
 - professional
 - Patents
 - management
- Business
 - Is not related to DBIS or CS

Doing Research

What Makes it High Quality?

Originality

- Is required
- Not necessarily comparable with relativity theory

Technical content

- Deep insight is more important than broad coverage

Presentation

- Typically a is a weakness in Russian research
- Often is the main reason for rejection

References

- Credits to other researchers

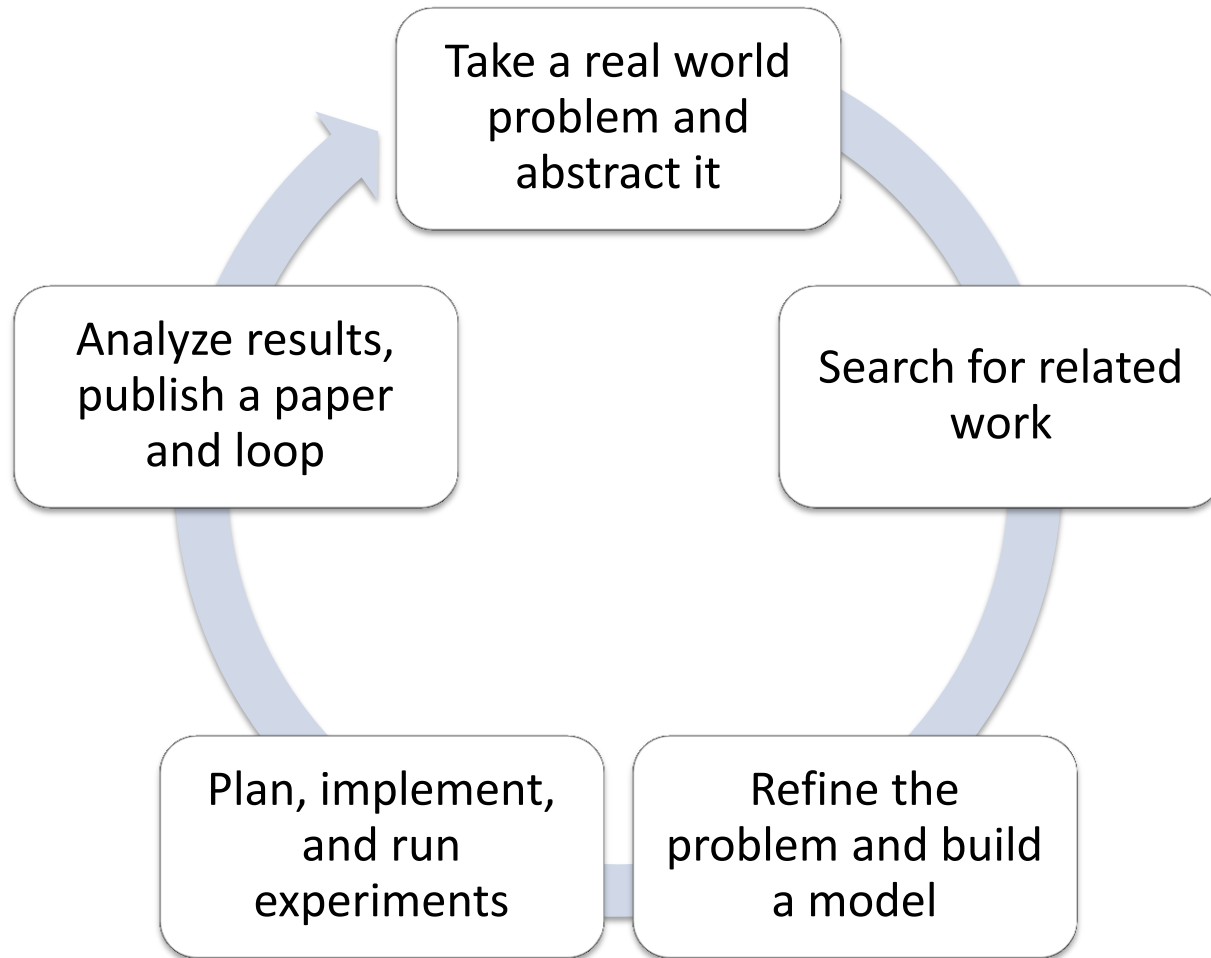
Reproducibility

- Converts the content into knowledge

Timeliness

- “Researches are always late” should be an exaggeration

Research Cycle



Identify and Shape the Problem

Meet requirements

- You should like it
- The problem Should be challenging and hard enough for your project (e.g. PhD project)
- The goals should be feasible

Consult

- Recent research publications
- Mass media
- Your advisor

Decide

- Do not rely completely on anyone, this is going to be your problem

Abstracting a Problem

1. Start with right real-world problem
2. Real-world problems are too complex for exhaustive study
3. Build a model
 - a. Cut out everything unimportant
 - b. Keep everything essential
4. Is the problem still there?
5. Iterate if needed



Abstraction Example



Real life object



The Model

Dive into Related Work

- Prove the novelty of your approach
- Cannot be exhaustive
- High quality is essential
- If you miss something important, it is your fault, and it is severe
- Highlight the differences in your approach, but be relevant



The Theoretical Model (Theory)

- ❑ Represents the knowledge gained from your research
- ❑ Provides the base for analysis and interpretation of experimental results
- ❑ Identifies the scope and applicability
- ❑ Makes the work reproducible
 - Provides guidelines for several contexts



Planning the Experiments

Do you really need them?

- You are expected to have
- E.F. Codd did not need them

Synthetic vs. real data

- Algorithms
- Human perception

What to measure?

- Effectiveness (e.g. precision and recall)
- Efficiency (e.g. response time)

Statistics

- Have you enough data?
- How reliable are measurements?

Presentation

Follow recommendations with caution

- What is motivating example?
- Explain your abstract model and concepts
- Prototype description only does not work
- Examples are essential
- Explain-by-example only does not work

Do it in your own way

- Be creative
- Dependencies: include only the stuff needed for main result
- Rehearse and try to present your work often

Publish Results

Quality measures

- Citations
- Reviewing
- Journal vs. conference
 - Acceptance rate
 - Revisions
- Community vs. formal measures

Ahead of Time

- Tycoon
Migrating persistent threads in a distributed object database system (early 90-ies)
Today: Virtual machines on a cloud
- Real-time distributed persistent object store (mid-90-ies)
Today: Key-value stores
- More examples can be found

Making It Work

Research Objectives

Organize already known facts scattered in several papers into a theory

- Essential for teaching
- The only knowledge which have chances to survive for centuries

Find solutions for challenging problems of real world

- The only way to real-world applications

The Paths

Basic research in an academic environment

Industrial projects, still in an academic environment

Trying to do it yourself: Start-Up

Long-Term Funding

Essential for structuring of the knowledge

Important for long projects (like PhD)

Needed for really deep insights

Provides stability (not always a plus)

Tends to go too far from real life problems

Industrial Projects

Typically short-term

Solving problems, rather than organizing knowledge

High risks are acceptable

Patents vs. publications

Do it on Your Own: Start-Up

Have a great idea, just be brave to start

What happens then?

- Thousands start-up companies are created every year, but the total remains approximately same
- At least 90% disappear during the first year
- Some continue as small-to-medium businesses
- Successful are acquired by big corporations
- Outliers are well-known, never try to repeat, create new market

Gaining the Popularity

What is Really Important for Public Success?

- Highlight Innovations
- Present innovations as improvements
- KISS
- Claim you are dropping some old stuff
- Never try to be precise or just completely correct
- To be successful, you have to blame something or at least claim it is outdated

Big Claims

The Claim	Translates to
Finally B-trees are outdated	Actually we are using somewhat different trees
Get rid of stupid database joins	Replace optimized database queries with nested loops implemented in the application code
Performance is not an issue	The implementation is not yet available
Flexible design	Real requirements are not known
Scalable architecture	Performance is not acceptable but hardware is not expensive

Refer to Principles

- From a book on software engineering: “The object paradigm is based on the principles of software engineering, and the relational paradigm, on mathematical principles.”
- Compare to aircrafts design based on psychology **INSTEAD OF** mathematics

Aircrafts



Mess Up the Logic

1. “All databases suffer from XXX problem”
2. Proof: Consider any DBMS, for example, xxSQL. It definitely suffers from XXX and this feature is annoying.
3. **Consequently, ANY** database has the same problem.

Quality Metrics

- “About 12000 persons already downloaded our product, consequently, there is something good in it.”
- Probably the product is really good, but the argument is NOT.
- Much more people are, unfortunately, addicted to illegal drugs.

Conclusions

- Research
 - Is hard, king's paths are not available
 - Provides perfect fulfillment
 - Is not a way to richness
- The quality of research does matter
- Identify your objectives and choose an appropriate path and model of success
- If you are in this room, you already decided

Good luck to all of you!