EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The quality of social simulation: an example from research policy modelling

Petra Ahrweiler
EA European Academy
Germany

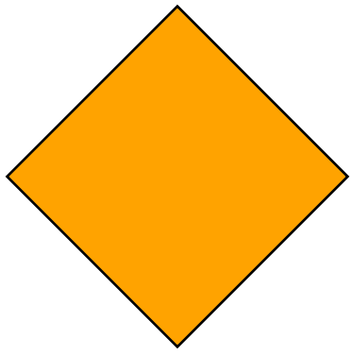
Nigel Gilbert
CRESS
University of Surrey, UK

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

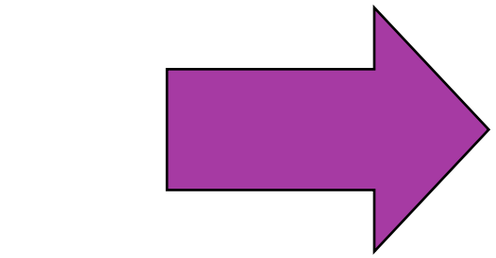
UNIVERSITY OF SURREY

The issue of quality we will talk about

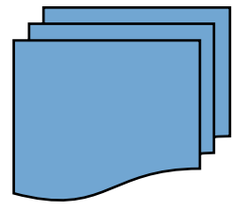


Model

Simulation



Simulation data



Abstraction

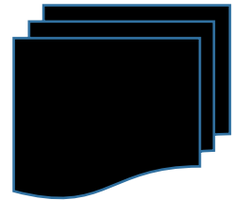
Similarity

=Quality?

Target

Data gathering

Collected data



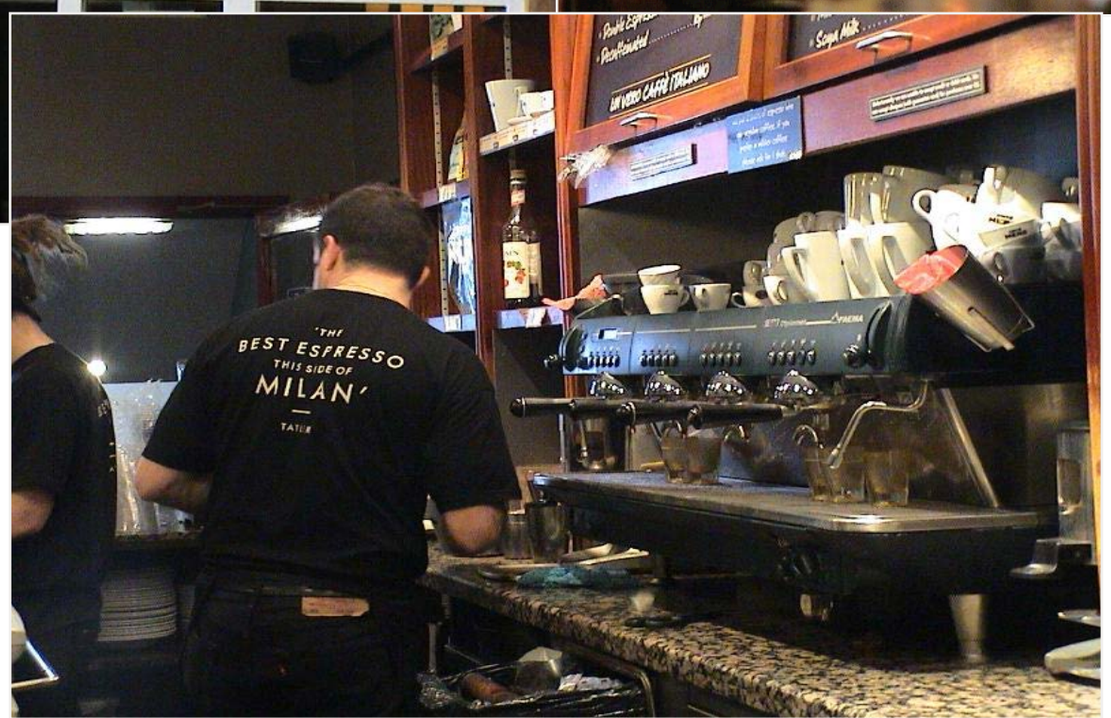
Centre for Research in **Social Simulation**

Caffarel
LA NUOVA LINEA
Fresh

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The simulation: Guildford's Caffè Nero



cress
Centre for Research in **Social Simulation**

The Standard View

- Verification
- Validation
  - Do the outputs for given inputs/parameters resemble observations of the target, although (because the processes being modelled are stochastic and because of unmeasured factors) identical outputs are not to be expected?
  - relies on a realist perspective because it refers to the observability of reality in order to compare the ’real‘ with artificial data produced by the simulation

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Caffè Nero

- Target:
  - Venetian Café
- Goal:
  - Getting “the atmosphere”(customers) and some profit (owners) from it
- Model:
  - by reducing the characteristics of the Venetian Café to a few parameters
- Measuring quality:
  - does the coffee taste the same as in Venice?
    - blind tasting
  - Is the noise level the same?
    - use a dB meter

cress
Centre for Research in **Social Simulation**

The Standard View

- Verification
- Validation
  - Do the outputs for given inputs/parameters resemble observations of the target, although (because the processes being modelled are stochastic and because of unmeasured factors) identical outputs are not to be expected?
  - relies on a realist perspective because it refers to the observability of reality in order to compare the ’real‘ with artificial data produced by the simulation
- Problems
  - Underdetermination: theories are under-determined by observational data or experience the same empirical data may be in accord with many alternative theories
  - Theory-ladeness of observations: Observations are supposed to validate theories, but in fact theories guide our observations, decide on our set of observables and prepare our interpretation of the data. At the very base of theory is again theory. The attempt to validate our theories by “pure” theory-neutral observational concepts is mistaken from the beginning.

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Is Caffè Nero a good simulation?

- Different concepts of the target lead to a different set of observables (either quantitative or qualitative)
- Important features of the concepts might not be observable at all

"A Venetian Café is a quiet place for reading my newspaper and relaxing with a good cup of coffee" (Nigel)

Observables: noise level (expectation: low), number of newspaper readers (expectation: high)

"A Venetian Café is a lively place to meet and talk to people with a good cup of coffee" (Petra)

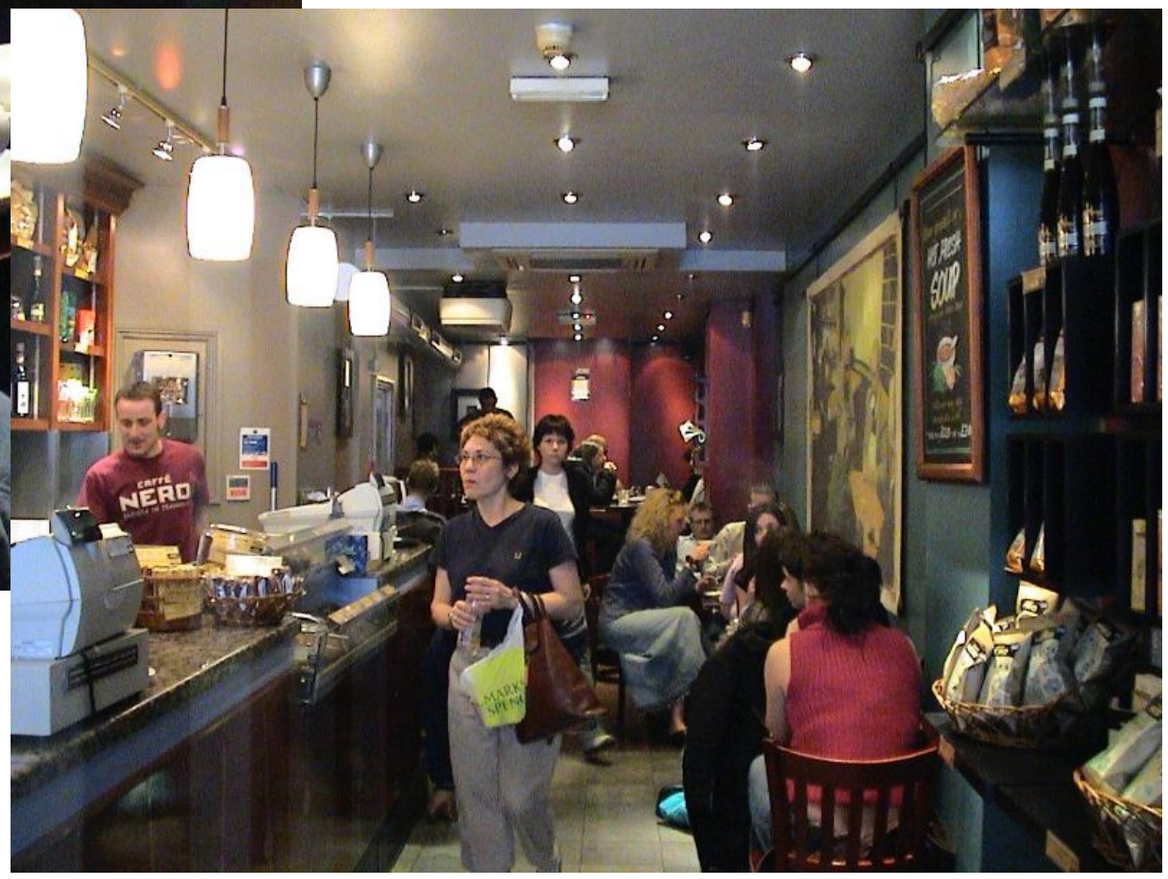
Observables: noise level (expectation: high), number of people talking (expectation: high)

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Nigel



Petra

cress
Centre for Research in **Social Simulation**

While these problems do not refute the standard view in principle but only emphasises difficulties in execution, the former arguments reveal problems arising from the logic of validity assessment.

We can try to marginalise, neglect or even deny these problems, but this will disclose our position as mere "believers" of the standard view.

The Constructivist View

- What you are comparing is not "the real world" and the simulation output; it is comparing *what you observe as the real world* with the output.
- Both are your constructions:
  - Your observations are dependent on your what you see as the relevant agents and their attributes
  - So is your simulation.
- They are just two ways of seeing the world.

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

However...

- Problem: „Anything goes“!?!
  - there seems to be no way to distinguish between different constructions/simulations in terms of “truth”, “objectivity”, “validity” etc. Science is going coffeehouse: everything is just construction, rhetorics and arbitrary talk. Can we so easily dismiss the possibility of evaluation?
- What about validation? What about assessing quality? What about checking against „reality“ (is there any)?

- To say there „is“ (sic!) construction, implies that there is something “real“ out there: namely the modellers, his or her constructions, and a “something”, which they refer to.
- At the base of the constructivist view is a reference to reality (how strange!)

cress
Centre for Research in **Social Simulation**

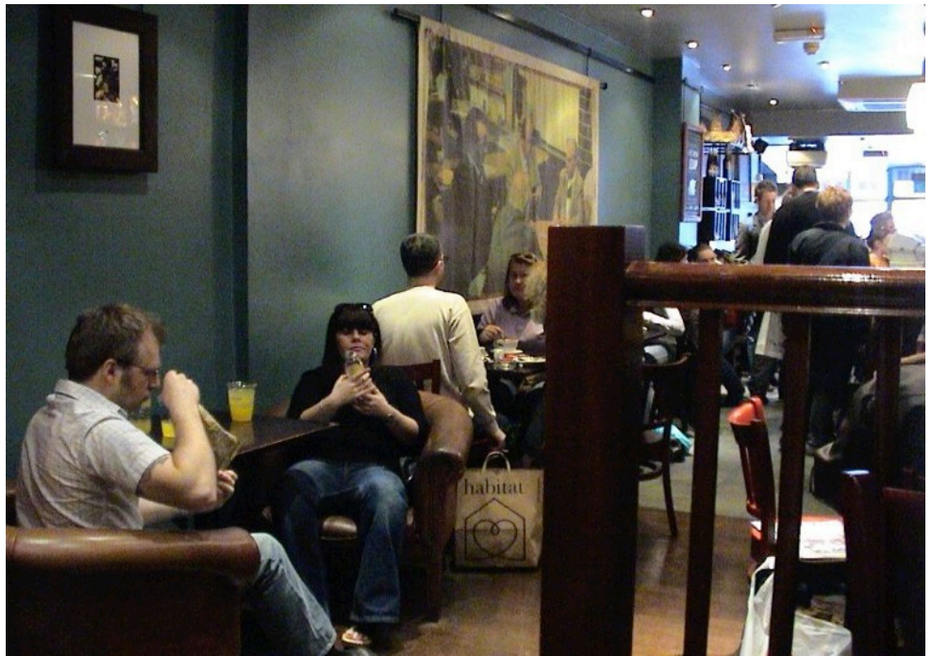
EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The User Community View

- So how can this problem of evaluation be resolved, if we have no direct access to an external empirical world?

- To find the answer, we have to remember that science is based on consensus: we agree about what is true (and what is true is what we agree about)



cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Conventions

- In order to evaluate a model, we need to know whose construction of the target it is being evaluated against
- But there is not complete freedom to construct anything you like
- At the base there are conventions and expectations which are *socially created and enforced*
- And therefore you can refer to these conventions to evaluate the quality of a model

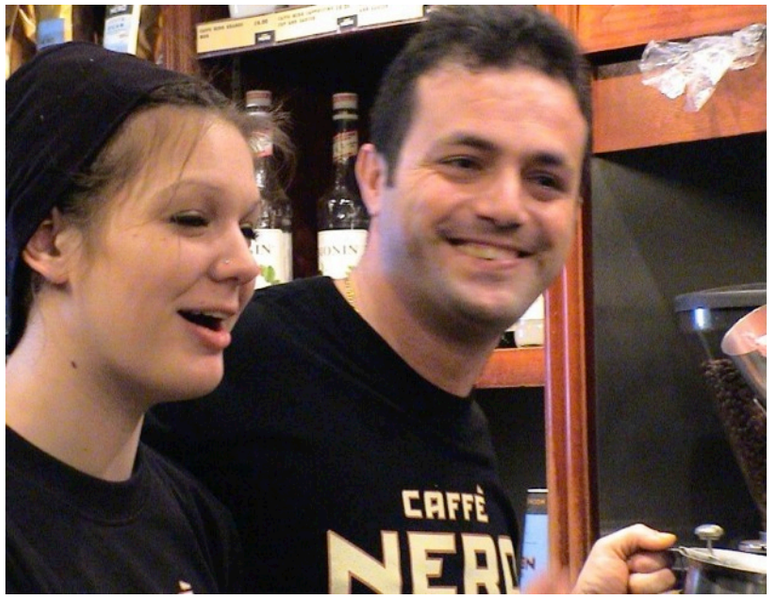
cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Evaluating Caffè Nero

We need a method which is based on the expectations, anticipations and experience of the community that uses it – for practical purposes, for intellectual understanding and for building new knowledge.



cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Evaluating social simulations

For computational models, we already have a social method of evaluation, in the ordinary (but sophisticated) institutions of (social) science and its practice

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Science evaluates itself

- The actual evaluation of science comes from answers to questions such as:
  - Do others accept the results as being coherent with existing knowledge?
  - Do others use it to support their work?
  - Do others use it to inspire their own investigations?
- The evaluation of scientific models comes from the practical evaluations of *users*, both scientists and others

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

An example: INFSO-SKIN

Centre for Research in **Social Simulation**

INFSO-SKIN

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

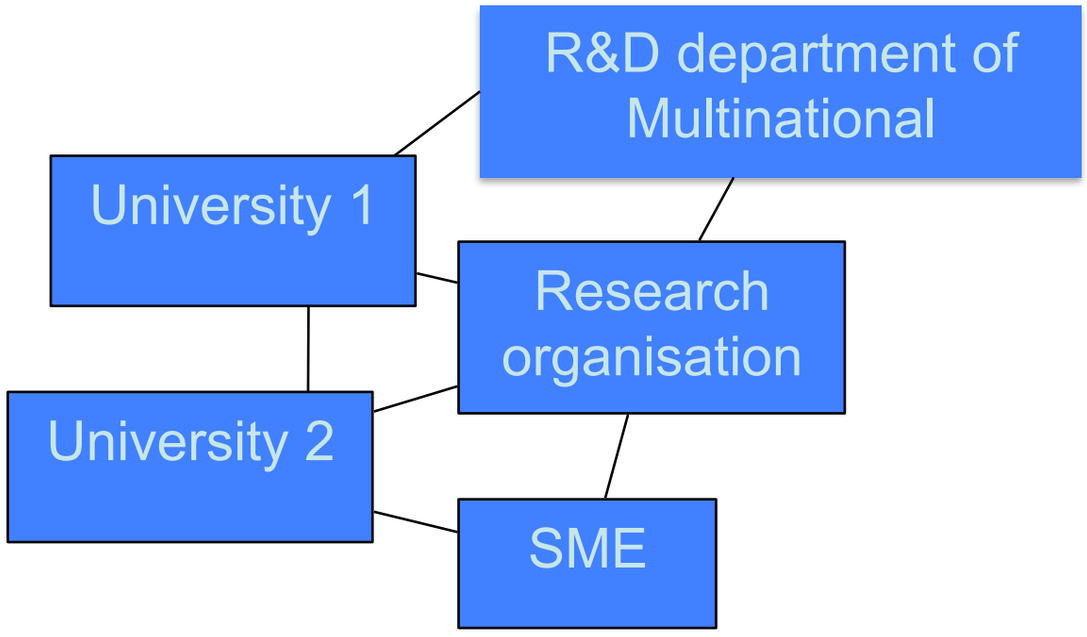
UNIVERSITY OF SURREY

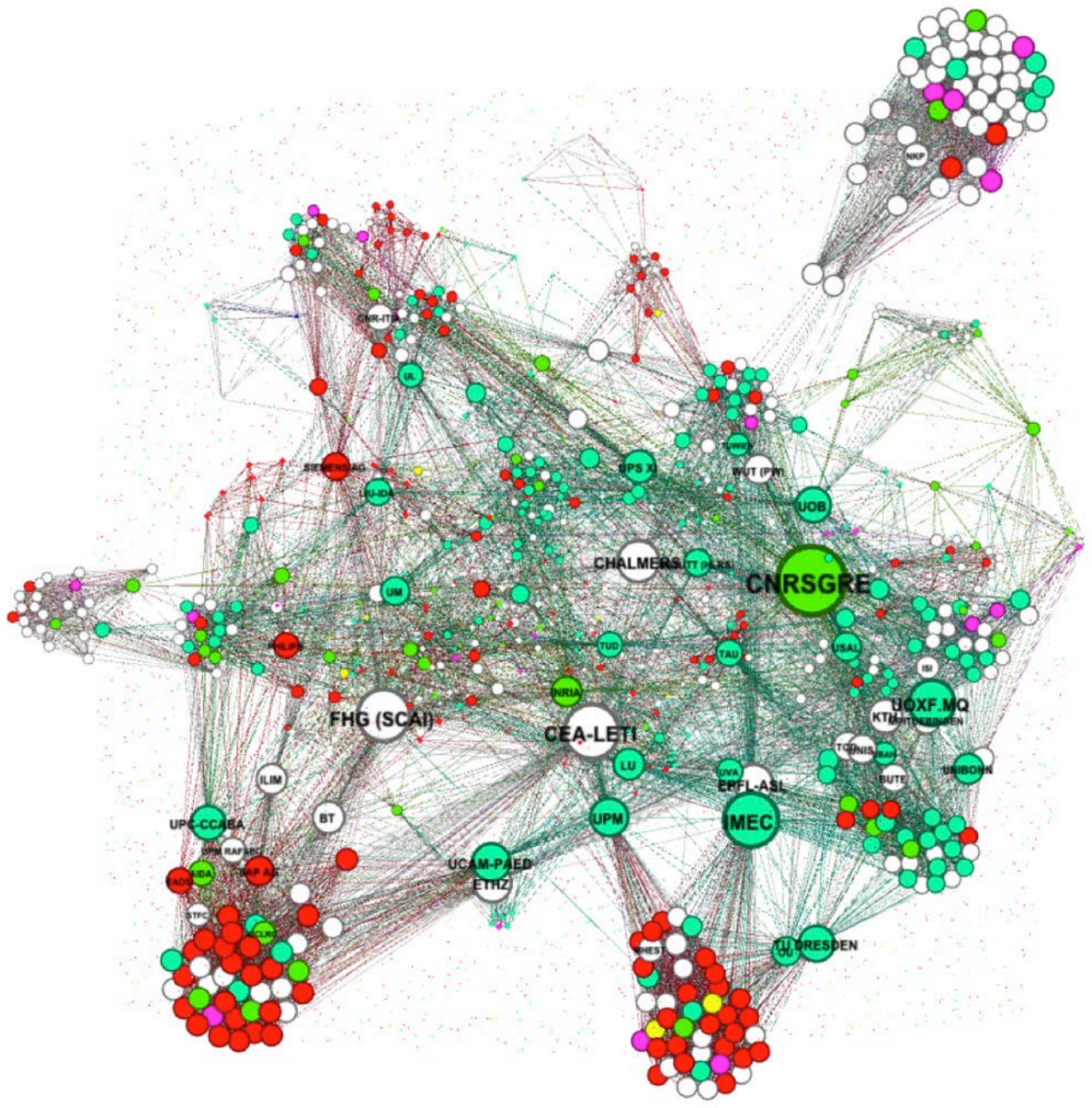
Research networks

EU FP6
research network in ICT (CA)

10.3

EU project consortium

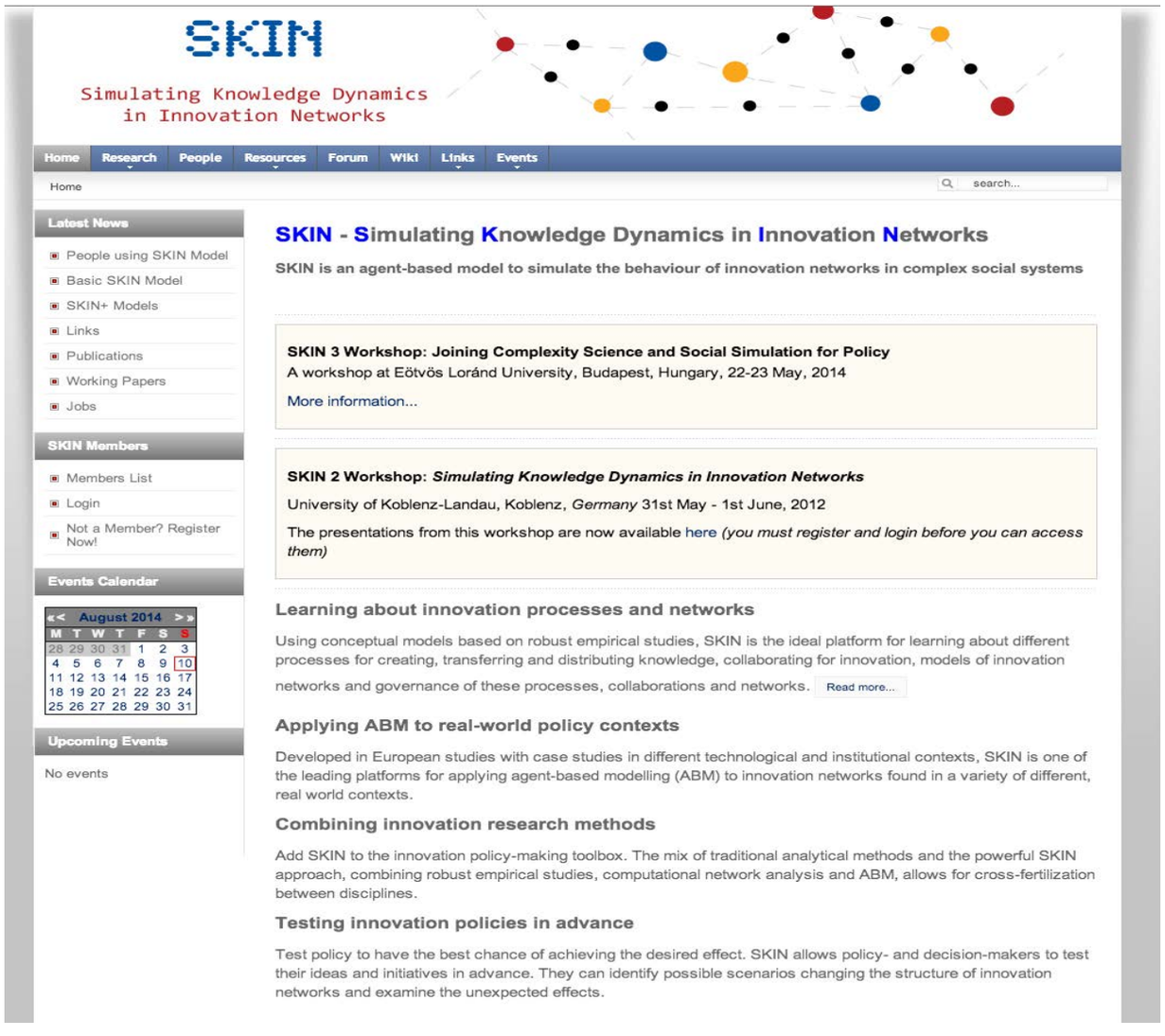




EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The SKIN model



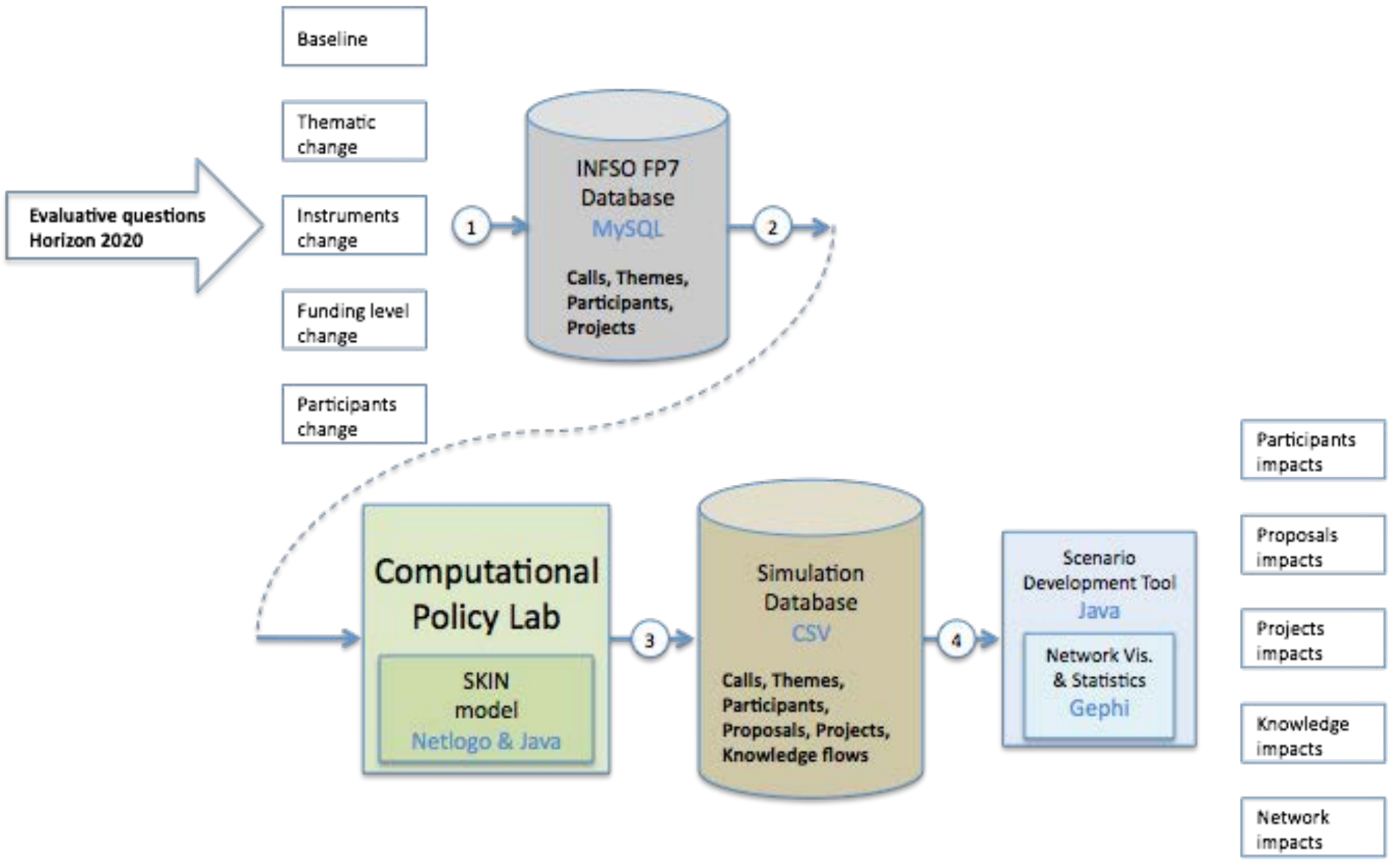
http://cress.soc.surrey.ac.uk/SKIN/

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The study workflow



cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Evaluative questions for Horizon 2020

- What if there are no changes?
- What if there are changes to the thematic areas?
- What if there are changes to the instruments of funding?
- What if there are interventions concerning the scope or outreach of funding?
- What if there are interventions concerning the participation of certain actors in the network (e.g. SMEs)?

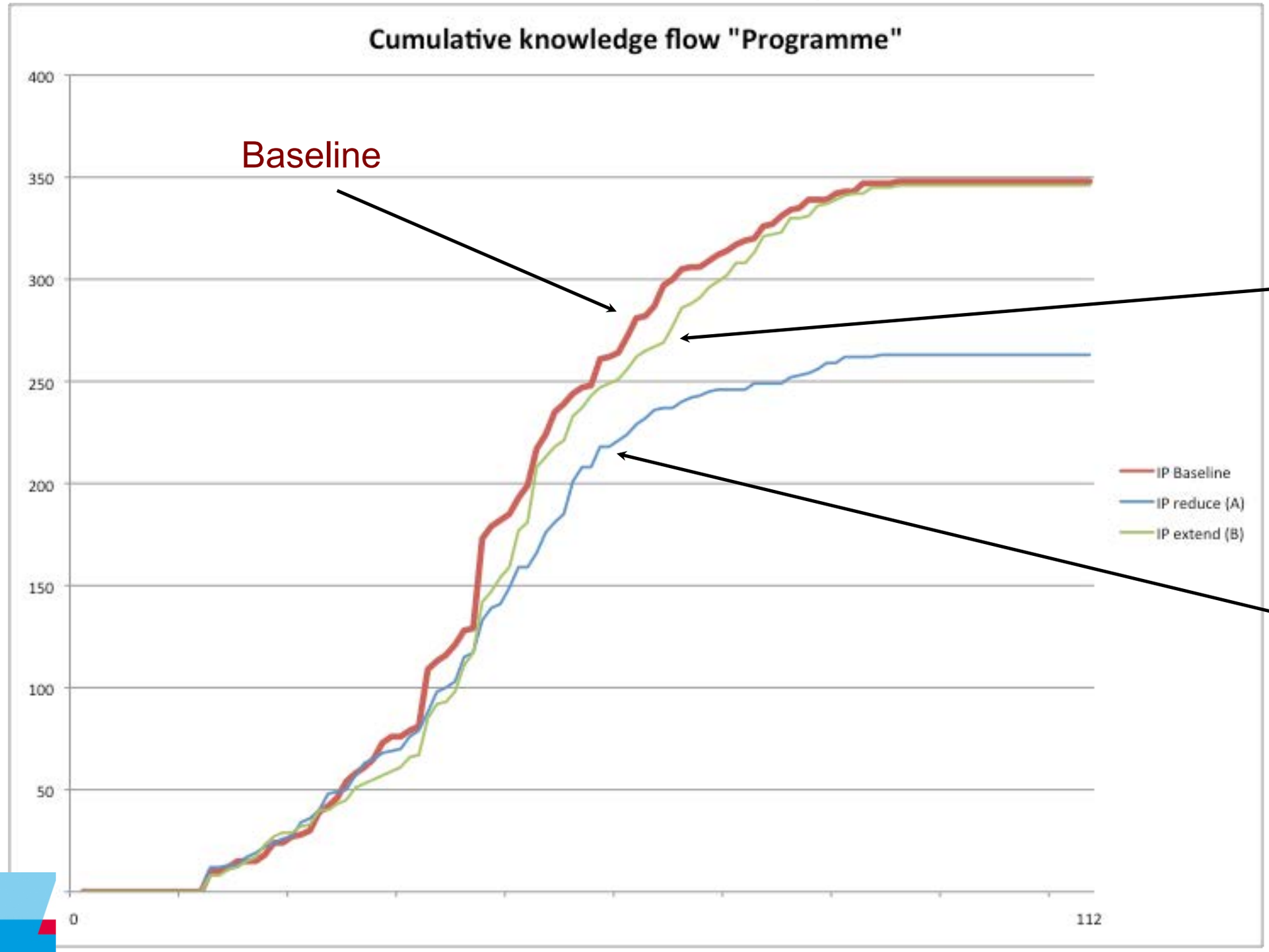
European Commission
Information Society and Media

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Results for

What-if we reduce/extend the number of funded themes?



cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The INFSO-SKIN model as seen by the Standard View

- Verification (+)
- Validation
  - relies on a realist perspective because it refers to the observability of reality in order to compare the 'real' with artificial data produced by the simulation
  - For addressing the evaluative questions of the stakeholders, we needed to create a simulation resembling their own world as observed as "empirical reality
  - The simulation needed to create the effect of similar complexity, similar structures and processes, and similar objects and options for interventions
  - To be under this *similarity threshold* would have led to the rejection of the model as a "toy model" that is not realistic and is under-determined by empirical data

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The INFSO-SKIN model as seen by the Standard View

In the eyes of the stakeholders, the more features of the model can be fed with and validated against empirical data points the better. Of course, there will be always an empirical "under-determination" of the model due to the necessary selection and abstraction process of model construction, empirical un-observables, missing data for observables, random features of the model and so on. However, to find the "right" trade-off between empirical under-determination and model credibility was a crucial issue in the discussions between the study team and the stakeholders.

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The INFSO-SKIN model as seen by the Constructivist View

- The strength of the modelling methodology lies in the opportunity to ask what-if questions (ex-ante evaluation), an option that is normally not easily available in the policy-making world
- INFSO-SKIN uses scenario modelling as a worksite for ‘reality constructions’

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

THE USER COMMUNITY VIEW IS THE MOST PROMISING, AND IN OUR EYES, THE MOST WORK-INTENSIVE MECHANISM TO ASSESS THE QUALITY OF THIS POLICY MODELLING EXERCISE.

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Identifying and negotiating the policy questions

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Negotiating the policy questions

- The Tender specification described the intended questions in detail, but…
- The stakeholder group (the 'clients')
  - worked out the meaning of these questions while they talked to us
  - dismissed the Tender questions and negotiated amongst each other for an alternative set
  - disagreed amongst themselves about which questions should be included, and about the priority of those included
  - did not fully understand the limitations of the methodology

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The process to get us there…

- Scan written project specification by client (in this case the Tender Specifications of DG INFSO) and identify the original set of questions
- Do literature review and context analysis for each question (policy background, scope, meaning etc.) to inform study team
- Meet stakeholders to get their views on written project specifications and their view on context of questions; inform the stakeholders about hwat your model is about, what it can and cannot do; discuss until stakeholder group and study team is "on the same page"
- Evaluate meeting and revise original set of questions if necessary (probably an iterative process between study team and different stakeholders individually where study team acts as coordinator and mediator of the process)
- Meet stakeholders to discuss final set of questions, get written consent on this, and get their hypotheses concerning potential answers and potential ways to address the questions
- Evaluate meeting and develop experiments that are able to operationalise the hypotheses and address the questions
- Meet stakeholders and get their feedback and consent that experiments meet questions/ hypotheses
- Evaluate meeting and refine experiment set-up concerning final set of questions

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Getting their best: users need to provide data

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The process to get us there…

- Identify the rough type of data required for the study from the project specifications
- Estimate financial resources for data access in the proposal of project to stakeholders (this can sometimes happen in interaction with the funding body)
- After second meeting with stakeholders (see section 2.3.1), identify relevant data concerning variables to answer study questions and address/test hypotheses of section 2.3.1*
- Communicate exact data requirements to stakeholders who are usually experts on their own empirical data environment*
- Review existing data bases including the ones stakeholders might hold or can get access to*
- Meet stakeholders to discuss data issues; make them understand and agree on scope and limitation of data access*
- If needed and required by stakeholders, collect data
- Meet stakeholders to discuss final database
- Evaluate meeting and develop data-to-model procedures*

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

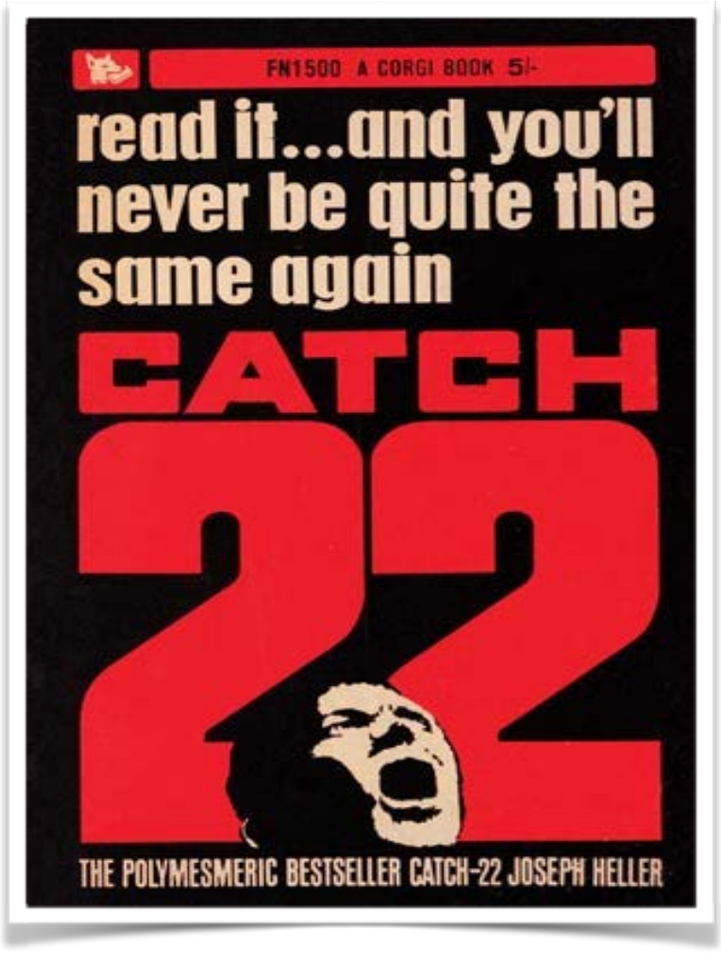
Negotiating the results

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

The modeller's Catch-22

- "The model reproduces what we already know"
  - why bother with a model?
- "The model predicts things we don't expect"
  - the model must be wrong



cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Negotiating the results

- For the stakeholders to trust the model (and its results), they needed to
  - understand the mechanisms represented in the model
  - feel that they have had an input into the design of the agent rules and characteristics
  - agree that the baseline simulations of FP7 were sufficiently close to what they observed had actually happened
  - be shown appealing visualisations and plots

- Then, they wanted ‘recommendations’, not ‘findings’
  - … more negotiation

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Conclusions

- To trust the quality of a simulation requires a trust in the quality of the *process* that produced its results.
- This process involves not just the model itself, but also the *interaction* between stakeholders and modellers
- So, modelling requires from both modellers and stakeholders
  - communication skills
  - patience
  - willingness to compromise
  - sufficient time
  - and motivation to ‘co-design’

cress
Centre for Research in **Social Simulation**

EA EUROPEAN ACADEMY
OF TECHNOLOGY AND INNOVATION ASSESSMENT

UNIVERSITY OF SURREY

Summary

How we should assess the quality of a model depends on:

- our assumptions about the world
  - an objective, external stable world?
  - a socially constructed perception of a world?
- the social context and the social conventions within which the model is designed, developed and assessed

cress
Centre for Research in **Social Simulation**

Thank you

petra.ahrweiler@ea-aw.de
n.gilbert@surrey.ac.uk