COMPLEX
Knowledge Based Climate Mitigation Systems for a Low Carbon Economy

Deliverable Number: D 6.12

Model Validation and Innovation

Date	**23/06/2016**
Report Number	
VERSION NUMBER:	0_2
Report Number:	Deliverable 6.12
Main Author:	Nick Winder, Isabelle Winder

Restricted pending possible publication

DIFFUSION LEVEL – RIP PENDING REVISION FOR INCLUSION IN FINAL REPORT AND PUBLICATION

PU	**PUBLIC**
RIP	**RESTRICTED INTERNAL AND PARTNERS**
RI	**RESTRICTED INTERNAL**
CO	**CONFIDENTIAL**

Author e-mail and contact details: nick.winder@ncl.ac.uk

CONTENTS

1. Introduction 1

2. Integrative Research and Stakeholder Engagement 2
   - 2.1 Integrative Research and Innovation 4
   - 2.2 Jonah's Paradox, Uncertainty and Meaninglessness 4
   - 2.3 Integrative Research and Reflexivity 7
   - 2.4 Rival Hypotheses or Deep Truth? 8

3. The Challenge: Model-Stakeholder Fusion 9
   - 3.1 Complex Causality and the Three Time Perspectives 10
   - 3.2 Management, Regulation and Innovation 11

4. Modelling the Modelling Process 12
   - 4.1 Example 1: Model-Stakeholder Fusion 14
   - 4.2 Example 2: Decision Support 16
   - 4.3 Distinguishing Mission-Creep from Possible Innovation 18

5. A Reflexive Experiment 18
   - 5.1 The Problem 20
   - 5.2 What were we hoping to achieve? 21
   - 5.3 How would we know we had succeeded? 22
   - 5.4 How did it go? 23

6. Conclusion; The COMPLEX approach to innovation management 24
   - 6.1 Phoenix-Cycle or Pre-Emptive Innovation? 26

1. Introduction

The COMPLEX project is a €5.4m effort funded to develop "Advanced Techno-economic Modelling Tools for Assessing the Costs and Impacts of (climate) Mitigation Policies". The call[1] contained the following words: "*Climate-energy-economy models are fundamental tools to evaluate* (sic) *mitigation strategies, assessing the costs and inform decision makers … currently available tools have relevant limitations such as the difficulty to represent pervasive technological developments, positive feedbacks, the difficulty to represent non-linearities, thresholds and irreversibility. Research should focus on the development and validation of new models, new model components or in the improvement/upgrading of existing models… The availability of large datasets for model validation purposes has to be taken into account, and their completeness should be improved. International collaboration to address the key challenges in Europe and globally is encouraged. The involvement of relevant stakeholders is highly recommended.*"

This combination of words strongly suggests a research focus on innovation, on situations in which small, weak causes can have unpredictable, large-scale effects, and on rapid system flips. An *integrative* approach was required that would bring natural scientists, human scientists and external stakeholders together to focus on key challenges in Europe. Although it was clearly intended that successful projects would devote substantial resources to developing and modifying computer models, there was also a theoretical dimension to the call, which invited applicants to consider the problem of validating models capable of representing technological developments, positive feedback, non-linearity and societal innovation.

It is relatively easy, in a multi-disciplinary community of modellers, to initiate unproductive arguments about the meanings of words like 'model' or 'reality'. In general, however, most modellers accept the distinction of a *process* model, designed and used to explore some body of theory abut a system of interest and a *decision-support* model, which is designed as a test-bed for hypotheses about how a system could or should be managed. Most would also accept that the distinction of process- from decision-support model is in the minds of those using it.

In conventional modelling practice, the word 'validation' describes the process of establishing that a decision-support model correctly simulates the historical behaviour of some system before using it as a platform for decision-support: you validate a decision-support model *ex post*, with the wisdom of hindsight and then use the decision-support model *ex ante* to predict future developments. In many modelling applications a distinction is made between 'calibration' (the parameterisation of equations or rule-sets that simulate dynamic processes) and 'validation', the process of establishing that goodness of fit between simulated and historical datasets is satisfactory. All models, whether they be of the decision-support or the process model type, must be calibrated. Decision-support models are often validated too and the call we respond to makes it clear that DG Research expect us to do this. COMPLEX has been very active in developing new databases and extending existing data resources for this purpose. However we have also addressed important questions about what the word 'validation' means in a world where scientists are modelling human activity systems.

Validation ties the model to a specific ontology and historical trajectory, effectively ignoring the histories that might have happened, but didn't. It is only scientifically defensible if a system is locally *time-symmetric*, i.e. if past system dynamics are a good guide to the future behaviour of the system on a given space-time scale. Time-symmetry was an implicit and unexamined assumption of policy-relevant research in the late 1990s and the early noughties of the 21st century, but suffered a knock in 2008 as financial institutions came close to collapse. Even as austerity measures were being implemented, there was a growing realisation that the EU needed modelling tools that could represent the unthinkable - catastrophic institutional and/or ecological change. That challenge requires systems modellers to design simulation models capable of representing rapid, possibly catastrophic system change in a realistic way. This is a non-trivial problem because the new human activity system will

[1] *ENV.2012.6.1-2*

depend critically on thoughts we have not yet thought and ideas we have yet to conceive. No amount of pre-existing data can fill that knowledge deficit in the present; we need to re-think the modelling process.

COMPLEX, now in its final year, was one of the projects selected for funding. One of its stated aims was to study the impact of rare but potentially catastrophic events, and try to get a handle on the probability that rational policies would have unforeseen and undesirable consequences. This research focus required us to assume socio-natural systems were time-asymmetric and that past system behaviour was often an unreliable guide to future dynamics. To build a simulation model that tracked past behaviours slavishly would require us to 'validate' it by ironing tricky, non-linear behaviours out. The result might be powerful tool for explaining historical dynamics, but would have little value as a tool for managing innovation. We need a new way of thinking about validation.

One of the most valuable heuristics of systems modelling is often summarised in the words '*if it is difficult, don't do it (yet)'*. The trick is to find a problem that has a similar look and feel, but is easier to solve. Solve that problem first and then look for ways of extending or generalising the solution method to cover the more difficult case. The problem addressed by this deliverable has two sources of difficulty; the first is that of building a mathematical or computer model capable of simulating time-asymmetries in a realistic way. The second source of difficulty is that of explaining how these models work without getting lost in mathematical formalism. We are not going to resolve those difficulties (yet). Rather, we are going to solve a simpler problem that is clearly related to our strategic task, that of negotiating a non-mathematical 'model' of a human activity system that can be used to manage system-flips and rapid innovations.

Readers may be interested to know that this deliverable is not the project's last word on the validation problem. COMPLEX convened a meeting on multi-model integration at IIASA in June 2016 at which this topic was discussed in some depth. Elena Rovenskaya, the leader of WP6, is preparing a report on this meeting that will be delivered within the next few weeks. We will not anticipate the results of that follow-up work here. Instead we will describe the process of building a 'model' of the COMPLEX project itself. The COMPLEX project is a human activity system. Its actions are constrained by the call quoted above. The project must be integrative, bringing hard scientists, human scientists and external stakeholders together; it must develop model infrastructure capable of representing rapid system transitions, work with pre-existing data and empirical evidence and address the validation problem. Finally it must facilitate the transition to a low carbon economy.

We cannot tackle this simpler problem unless we understand it, and understanding requires us to be clear about what we mean when we use words like 'model', 'innovation' or 'stakeholder'. We therefore apologise for the slightly pedantic style of the next two sections. Our aim is not so much to educate as to provide the level of semantic precision needed to construct a qualitative model of the COMPLEX project and use it to explore the validation problem.

2. Integrative Research and Stakeholder Engagement

The story is told of a climate scientist who visited the Arctic to meet with key stakeholders - people working in primary exploitative industries like mining, fishing and herding which would be influenced by changing climate. In one of these northern towns citizens listened attentively while he explained about small global temperature rises, their impact on atmospheric and ocean circulation, ice-melt, sea-level rise and extreme weather events. At the end of his talk he invited the audience to help him understand what the impact of these changes would be on their own lives and was met with polite, but confused silence. He was a good listener and encouraged them to talk about the problems that most engaged their own community. They told him their biggest problem was the lack of young women, who tended to go south for education and not come back. The sex ratio was hopelessly imbalanced and young men had trouble finding mates.

The process of de-population, it turned out, was subject to a positive feedback loop. The words 'positive' and 'negative' should not be taken to mean 'good' or 'bad'; many positive feedback loops are vicious. A negative feedback loop occurs when a control mechanism reverses a trend, as a thermostat switches off a heater to prevent a room getting too warm. By analogy, a positive feedback loop reinforces a trend.

As the population of this Arctic town dwindled, the cost of maintaining good ante-natal and paediatric care was becoming harder to meet. It seemed likely, in the near future, that women whose labour was not going well would have to be transported long distances to the nearest maternity hospital. This perceived risk was accelerating out-migration among young women. Moreover, the parents of young boys were themselves moving to regions where the sex ratio was more even, further accelerating de-population. Even opportunities for shopping seemed to accelerate depopulation. Buying durable goods was not a problem, but perishable foodstuffs like fresh fruit and vegetables could only be bought if the cost of hauling it in and the benefits of trading were met. As the community shrank, there was a genuine fear that there would no longer be bananas in the supermarket.

These positive feedback loops were having a very negative effect on morale. Although some of the citizens were naturally interested in research on climate change, nothing they could do would change the way the climate system evolved - they would have to adapt to climate change or migrate like everyone else. Consequently, they did not see themselves as stakeholders in any research project on climate change. A research project on the demography and economy of Arctic settlements would have been a different proposition. They could see themselves becoming stakeholders in that.

This salutary tale illustrates the importance of thinking carefully about what words like ‘stakeholder’ signify. The COMPLEX project, for example, could easily write about ‘stakeholders’ in hand-waving terms; talking about ‘the general public’, say, or the energy sector, but to do so would be to exaggerate our own agency and domain of influence. If we are to work in an integrative way, bringing scientists and external stakeholders together to create new knowledge, we need a definition of the word ‘stakeholder’ that reconciles our research interests to their concerns.

A common-sense definition would be that a stakeholder is an organism, individual, population, community, institution or ecosystem that ‘has a stake’ in some thing. That phrase, ‘organism, individual, population, community, institution or ecosystem’ could reasonably be replaced with the indicative term ‘system’ and the definition re-emerges as: a ‘stakeholder’ is a system that ‘has a stake’ in some thing. The phrase ‘has a stake’ gives the misleading impression of a conscious and deliberate investment. Many stakeholders are unwilling and unacknowledged; wild orang-utans, for example, are unwilling and unacknowledged stakeholders in forest clearance projects. In practice, stakeholders tend to be systems whose dynamics are influenced by the thing in which they ‘have a stake’. The dynamics of those orang populations, for example, are influenced by forest clearance projects.

Finally we need to consider what the words ‘some thing’ signify and it is enough to consider the types of ‘things’ in which one might be a stakeholder. Clearly, the banking system is a ‘thing’ and so is the COMPLEX project, a lottery would qualify and so too would a business. All these ‘things’ are human activity systems of one sort or another.

If we put all these ideas together we come up with the following:

if the dynamics of some system Y have been modified by some named human activity system X, then Y can be described as a *stakeholder* in X.

This definition has two significant implications for the case-study described here: First, it implies that human activity system X is a stakeholder in itself. For example, all the scientists working on the COMPLEX project, together with partner institutions and funding agencies, are also stakeholders in the COMPLEX project. So too are the NGOs and CSOs with which we consult, the policy makers and energy companies we work with and those of our peers who take the time to read our findings. The size and character of these stakeholdings may vary, but there is no qualitative difference between a scientist’s stakeholding in the COMPLEX project and that of the people who attend our meetings or work with us on a daily basis. Second, it implies that many of the people, institutions and ecosystems we would dearly like to influence are not COMPLEX stakeholders. Our project has had no measurable impact on the ‘general public’, the Great Barrier Reef, national governments, global warming, or the coal industry, for example - the COMPLEX project has not re-shaped system dynamics on these scales.

Like most research projects, the strongest stakeholders in COMPLEX are the scientists themselves, their employers and funders. This group of internal stakeholders is followed, at some distance, by small groups of external stakeholders that the project interacts with regularly. Readers interested to know who or what those stakeholder-systems are can consult Volume I of our final scientific report, which will be available on-line at the project's web page in September 2016.

2.1 Integrative Research and Innovation

The old distinction of policy makers from scientific experts and stakeholders is the first casualty of the integrative approach, which moves the focus of research from doing science for external stakeholders to doing science with them. Over the last two decades this focus on doing things with external stakeholders has come to be described as 'co-production', a concept often associated with the Nobel laureate Elinor Ostrom. Indeed, in recent years 'co-production' has become a catch-all term that describes a great range of methods and traditions. Here we will use the word 'integrative' to describe the general case and preserve Ostrom's term for that specific intellectual tradition[2].

Epistemic diversity plays a role in integrative research comparable to that of biodiversity in ecology; it is a source of adaptive potential and ecodynamic resilience. Many natural scientists find this idea counter-intuitive and even unscientific: surely, if two people have logically irreconcilable worldviews, then at least one of them - possibly both - must be wrong? To pretend otherwise would be to slip into relativism - the error of believing that that any viewpoint, however absurd or baseless, must be taken equally seriously. In practice, however, integrative research is seldom relativistic. Rather, it exploits the distinction, attributed to Nils Bohr[3], of a *deep truth* from a trivial truth. To contradict a trivial truth would be to speak a trivial falsehood, but it is sometimes possible to contradict a deep truth and speak another deep truth. Integrative research deals with deep truths.

Integrative research often creates opportunities for *innovation* - for enabling humans to change the course of history by changing their minds. Innovation is arguably the single most significant difference between science in the second half of the 20th century and that which went before. Pre-WWII science often dealt with hypothesis testing, and philosophers argued about the relative merits of verification and falsification. In post-war science, however, predictions are commonly used to trigger or encourage innovations that would make their truth or falsity undecidable. The COMPLEX project, for example, does not strive to demonstrate the truth or falsity of climate-change predictions; it seeks innovative solutions to socio-economic problems that will modify human behaviour in a way that renders the truth or falsity of those predictions undecidable.

Although integrative science is a product of the later 20th century, ideas about logically undecidable predictions are not. They are clearly described in Aristotle's *Posterior Analytics*, for example, and have been discussed under the rubric of 'Future Contingents[4]' for many centuries.

2.2 Jonah's Paradox, Uncertainty and Meaninglessness

Elsewhere we have written of two related ideas: Jonah's Law[5] and Jonah's Paradox. Jonah's Law states that *humans can only predict the course of history in respect of phenomena they cannot*

---

[2] Ostrum, E (1990) Governing the Commons

[3] In: Discussions with Einstein on epistemological problems in quantum physics. Bohr writes of the old saying about 'the two kinds of truth. To the one kind belong statements so simple and clear that the opposite assertion obviously could not be defended. The other kind, the so-called "deep truths," are statements in which the opposite also contains deep truth.'

[4] A Future Contingent is a statement of the form: if it rains, there will be a sea-battle tomorrow. If it rains, we will be able to establish its truth or falsity ex post. If it does not rain, we will never know whether the prediction was true. Predictions in the integrative sciences are usually contingent on formal models of natural systems and current knowledge about human-environment interaction. If humans innovate in ways that change or refute these beliefs, the truth or falsity of those predictions will never be known.

[5] Jonah is said to have refused a direct instruction from god to prophesy the destruction of Nineveh. His grounds were that the iniquitous population would innovate - making reparation so that god would forgive them. This

influence and can only change the course of history in circumstances where their predictions are potentially meaningless. Jonah's Paradox (see Box 1) states that *any attempt to predict the future of an innovative system, even a prediction based on a perfect understanding of system dynamics and initial conditions, will generate logically undecidable propositions if the prediction is shared*.

Box 1: Jonah's Innovation Paradox

Professor Jonah, the distinguished political scientist, has built a model which predicts that Hillary Trump will win the next election because of tactical voting. Jonah is absolutely correct (the tactical voters really are going to put Trump into office) and everyone knows Jonah is a brilliant, indeed, omniscient modeller. Jonah doesn't favour H.T. but, as one voter among many, cannot change the outcome of an election. So he takes out an advertisement in the newspaper and communicates his fears to the population. You are a potential tactical voter. When you read the advert, you must decide how to respond. Your first question might be: is Professor Jonah right or wrong?

You know that Jonah can predict the result of the election and that he is omniscient. It might seem logical to vote as if he were right. Unfortunately, you also know that he has shared his knowledge with many others. If enough of these electors change their voting behaviour, then Jonah's prediction (reputation notwithstanding) will be wrong and you should vote accordingly. However, if enough people disregard Professor Jonah, he will have been right and you should act as if this were so, . . .

As soon as Jonah tells everyone what he knows with absolute certainty, he generates a potentially undecidable proposition. You have no basis to decide whether the assertion is true or false *ex ante* despite the fact that you know he is omniscient and so was undoubtedly correct at the time he wrote the advertisement. You simply have to resort to guesswork or wait and see how it all turns out. Of course, if Jonah hadn't taken out that advert, the truth of the prediction would have been ensured, but then H.T. would win the election and Jonah would have failed to change the course of history.

This is Jonah's Innovation Paradox.

Since integrative science is intended to facilitate, or at least understand, innovation and innovations create new knowledge that influences human behaviour and change system dynamics, integrative researchers need to be aware of Jonah's Law and Jonah's paradox. Our predictions are always uncertain and sometimes downright meaningless.

Consider, for example, a prediction of the mean summer temperature on the top of Ben Nevis in the year 2500. Categories like 'mountains' and 'temperatures' are ontologically robust. Those making the prediction could even use probability methods to get a handle on the level of uncertainty because the ontology is robust enough to make it possible to speak of hypothetical populations of mountains and long-term climate scenaria. Predicting the gross domestic product of Scotland in 2500, however, would imply that the polity we call 'Scotland' and its economy would still be recognizable 500 years from now. Such a prediction is not merely uncertain; it is meaningless because a host of innovations could intervene which might sweep our geo-political worldview aside. It is not meaningful to speak abstractly about populations of polities a little like Scotland 500 years from now because human activity systems can innovate[6].

The distinction of uncertainty from meaninglessness is a rich source of misunderstanding between the analytical sciences and the discursive humanities and much of that misunderstanding comes to rest on the concept of 'reality'. In study domains like cosmology or palaeontology, where human agency is

would falsify Jonah's prophecy. Jonah was so anxious to avoid false prophecy that he ran away to sea where a storm and passing fish forced his hand. When he finally conveyed god's message, it all turned out as anticipated: the people made reparation, Nineveh was saved and Jonah sulked for days.

[6] *One could argue pedantically that a catastrophic geological event might blow a hole in the Grampian mountains or that miners might be inspired to destroy Ben Nevis and that these events would render the other prediction meaningless too, but the mountain has existed since the Devonian period (400,000,000 years). The polity that is Scotland has changed a great deal in 4 centuries and Ben Nevis remained a mountain throughout that period. The probabilities of predictive meaninglessness differ by many orders of magnitude.*

limited and innovation seems unlikely, it may be reasonable to speak of ‘reality’ as that which is independent of human knowledge and agency, but even in these disciplines there are at least two different flavours of ‘reality’ to consider. Material reality refers to objective, observable things like mountains and cloud-formations, while abstract reality refers to ideational structures, mathematics and symbolic reasoning.

The tension between rationalist and empiricist views of reality can make it difficult to integrate the research output of natural scientists trained in different traditions. Pure mathematicians and statisticians, for example, tend to have very different worldviews. The situation becomes even more complex when humanists enter the team, bringing with them ideas about ethics and consensus. When humanists speak about socially constructed ‘realities’, natural scientists often become uncomfortable and accuse them of solipsism - of believing that Pythagoras’ theorem or meteorites are figments of the imagination. This is as unhelpful as it is incorrect.

One of the great challenges for integrative science is to stop researchers getting locked into territorial displays about reality. A defensible strategy for avoiding these squabbles is to accept that reality is a *phenomenon*, a bundle of sense-data and ideas that have been shaped by individual experience, neurophysiology, embodied analytical skills and socially constructed consensus (knowledge). Different people can interpret the same sensory inputs in different ways. It is a matter of indifference whether you refer to those perceptual differences as ‘reality’ or ‘phenomena’ as long as you understand that they are always shaped by prior knowledge, and that knowledge really is socially constructed.

Innovations are time-asymmetric; they cannot be predicted *ex ante* because they are contingent on knowledge we do not yet possess. They can, of course, be explained, *ex post*, with the wisdom of hindsight. Time-asymmetry is one of the greatest methodological challenges to integrative research and, ironically, one of its principal strengths. A simulation model that predicts a catastrophic system collapse, for example, could act as a trigger for pre-emptive innovation, but it could equally well be dismissed on the grounds that, as we come closer to the predicted event, humans will innovate in a (currently unimagined) way that miraculously avoids the catastrophe. Jonah’s law states that innovation is possible and this report will try to describe ways of making innovation more likely, but a blind faith in the ability of future generations to solve problems we cannot tackle can have a stultifying effect on human activity systems.

2.3 Integrative Research and Reflexivity

Stakeholder systems have a fractal structure. The word *fractal* refers to an object in which the same type of pattern is repeated on every scale. A government agency, for example, is a named human activity system of organisms whose collective action is influenced by other stakeholders. The same can be said of the community of lobbyists that represents a commercial sector, the community of migratory wildfowl wintering in a marshland, or a group of birdwatchers. Each of these stakeholder communities has a characteristic space-time scale - a domain of influence, if you will.

Of course we are not saying that each stakeholder community is exactly the same as all the others. Some stakeholder communities have only one member, for example. Some stakeholder communities may be populations of micro-organisms. Some stakeholder communities are always co-located in space-time - an individual human would be a case in point - but many are almost never co-located. The fundamental idea is that all are stakeholders in the sense we have defined and each has a characteristic space-time signature.

We have tried to simplify the model we are developing by restricting the stakeholder concept to systems whose dynamics are shaped by ***named human activity systems***. The action of naming is significant because it implies that the modeller is aware of the existence of that system and acknowledges its existence. Every time the stakeholder chain reaches some non-human system - a population of animals or plants, say, or a system whose existence the modeller is unaware of - the chain of recursive calls to the class 'stakeholder' bottoms-out. So when we say that stakeholder systems have a fractal structure, what we are saying is that, even if we ignore recursive calls to the class 'stakeholder' that lead us into non-human systems, the chain of recursive calls to the class stakeholder never quite bottoms out.

Consider a football match, for example. Every player is a stakeholder in the game, in their own team and in the opposing team. The crowds in the stands, the referees, the linesmen, the film crews, the personnel manning the turnstiles, the police officers keeping order, the TV companies, agents, publicists, advertisers, … The list of stakeholders never bottoms out. The same organism can often belong to many stakeholder communities, each with a different space-time signature. An individual supporter, for example can be a football coach or an off-duty police officer, and can switch roles by wearing different stakeholder 'hats' in different contexts[7].

This role-switching behaviour is largely unconscious and individuals can flip between roles and culturally embedded habits in ways that allow them to accommodate logical and behavioural inconsistencies. The researcher dedicated to reducing carbon emissions, for example, may drive a car, take foreign holidays and be a member of a pension scheme that invests heavily in petrochemical industries. Such inconsistencies are often amplified on larger space-time scales, leading to problems of cross-policy compliance in which well-meaning, habitually embedded activities in one context frustrate policy initiatives in another.

An integrative research project like COMPLEX is a *reflexive* structure; the ontology of the research team reflects that of the study-domain on which it works. The research team is a stakeholder, a diverse community of individuals trying to influence the behaviour of diverse communities of (external) stakeholders, some of which are themselves trying to understand and influence diverse (external) stakeholder communities of their own. The result is a network of fractal systems (hereafter a *fractal network*), each of which has its own space-time signature and context. A wild orang-utan, for example, could be a stakeholder in a research project on the impact of forest clearance, so too are the trees she lives in, the researchers who work on the orangs, the farmers who want to work the cleared land, the snack manufacturer looking for cheap sources of palm oil, the committee that funded research and the politicians and civil servants who control research investment, and so on.

[7] *The popular literature on fractals is often self-contradictory, suggesting that fractals are infinitely self-repeating while arguing that material objects like coastlines and trees are really fractals. Trees and coastlines are not infinitely self-repeating.*

In general the further one travels along that fractal network, the weaker the stakeholding becomes. However, a seminal paper titled *the Strength of Weak Ties* by Mark Granovetter explains that long-distance, weak stakeholder links may be significant triggers of rapid systemic change. The stakeholders with which a project has closest links are often locked in to patterns of reciprocal obligation that prevent them taking much-needed action. Research projects on orang conservation, for example, may have less impact on politicians than voters and lobbyists. Politicians experience conflicts of interest that can make it impossible for scientists to influence them. These receptivity barriers become manifest as 'boundary constraints' that limit the agency of some stakeholders. Some of those boundary constraints are explicit, but many are unspoken and unacknowledged.

Powerful stakeholders do not like to be told they are solving the wrong problem and create a barrier of administrative and political blocking actions that prevent integrative research projects innovating in a disruptive or inconvenient way. This, in turn, generates a strong professional selection-pressure that differentially advances the careers of scientists who don't rock the boat too much. Institutional vetoes tend to remain in place as long as institutions are perceived as strong and credible and may even be tightened in difficult times, leading to social exclusion and even repression. The great innovation cascades occur when that hegemony slips, as it did in Western Europe in 1920s and again in the 1960s and in the non-aligned and Soviet blocs in the 1990s. These cascades create opportunities for mavericks with lots of weak, long-distance connections and fewer obligations and constraints.

The mavericks posterity credits with great foresight and those whom textbooks say played pivotal roles in innovation-cascades are usually part of a wider undercurrent of social change and re-conceptualisation that was suppressed by institutional vetoes. After the deluge, those mavericks are feted as innovators and institutions repair themselves by shifting the responsibility onto the old regime. This couldn't happen today, of course, because now we have scientific advisors who really understand how science works - advisors whose credentials have been vetted by powerful institutions. In this way a new set of institutional vetoes are created and the stick-slip cycle begins again.

2.4 Rival Hypotheses or Deep Truth?

Receptivity barriers do not belong to the domain of the natural sciences - they are social, cultural and economic artefacts that limit the scope for innovation. Stakeholders committed to innovation naturally find them frustrating and many are outspoken critics of powerful institutions; arguing, in effect, that the world is going to the dogs, but that everything would be alright if only institutional vetoes were relaxed and people listened to them.

Institutional actors, of course, see the process from a very different perspective. They believe eco-warriors and activists are extremists who fail to appreciate how much damage would be sustained if powerful institutions were to collapse. The phrase 'too big to fail', which was heard frequently in the aftermath of the 2008 crash, epitomises this institutional perspective.

One might imagine that an integrative project like COMPLEX would champion a *third way* approach - the revolutionary eco-warriors are extremists and so too are the ultra-conservative institutional actors, what we really need is a compromise position that lies somewhere between these extremes. However, as we explained in Section 2.1 above, COMPLEX is open to the possibility that there is no compromise position. Perhaps the eco-warrior and the apparatchik are not extremists at all, but members of different political 'disciplines' with different space-time signatures; perhaps each group is looking at some polarised 'deep truth' from a different perspective; perhaps causality is complex.

If so, the reason the reformer believes powerful institutions systematically suppress and veto uncomfortable innovations and undermine cultural and ecological life-support systems is that they do. The reason institutional actors think eco-warriors are disruptive iconoclasts willing to compromise intuitional stability in order to achieve their goals is that they are. The dialectic tension between these reciprocally coupled cause-effect systems, each with its own characteristic space-time signature, establishes dynamic equilibria that hold human activity systems in a familiar basin of attraction. Innovations occur when the balance between these polarised truths flickers or wobbles in a way that allows new ideas to infiltrate mainstream thought and new behaviour patterns to emerge.

If both perspectives are true, then the fears and preoccupations of both groups must be taken seriously too. The collapse of great institutions would indeed be catastrophic. Indeed, we have ample evidence that this is so; institutional collapses from the French revolution to the de-stabilisation of established regimes in Iraq, Libya and Syria have all had unwelcome knock-on effects. Predictions about environmental degradation, the limits to growth and the importance of subsidiarity which, 40 years ago, were the preserve of tree-huggers and the lunatic fringe of science are now widely accepted by mainstream scientists and policy makers. Revolution is not an option, and neither is business as usual.

Integrative science requires us to explore the possibility of switching control between these poles and allowing micro-scale events to trigger innovations that will change the dynamics of the system being studied. Our aim must be to bounce it from a dangerous attractor to one that is gentler and more sustainable. Of course research projects have limited influence. Powerful institutions only become stakeholders in integrative projects on their own terms and systematically ignore uncomfortable truths. The same can also be said of reformers and political activists.

Nobody said integrative research was easy.

3. The Challenge: Model-Stakeholder Fusion

Most of the models that currently reside in the COMPLEX Model Repository[8] fall into one of two broad categories. The first type either uses large datasets or contains agents that represent powerful institutions and are used to explore the likely consequences of large-scale policy options - should governments impose a carbon tax? Are austerity economics more eco-friendly than Keynesian stimulus? How should energy utilities handle intermittency problems? Alongside the models that speak to power, we have a second set of modelling tools aimed at regional or municipal level. Many of these small-scale models are designed for integrative work. The smaller models are not built by scientific experts for powerful stakeholders, but rather are negotiated by alliances of weaker stakeholders, many of whom want to reduce their dependence on energy utilities and find ways to work-round the constraints imposed on them by governments, politicians and lobbyists.

Both types of model were the products of a research initiative intended to attune our modelling efforts so finely to stakeholder perceptions that the result could reasonably be presented as ‘model-stakeholder fusion’. Our research proposal said that COMPLEX would not restrict itself to conventional modelling tools, but would operate in a ‘generalized socio-environmental model space’ that would include empirical models, conceptual models, complex computer simulations and data-sets.

Some members of the team, including the co-ordinator, expressed concerns about this aim at the proposal-writing stage. It was hard to see how we could characterise a ‘generalised socio-environmental model-space’ and negotiate a ‘model-stakeholder fusion’ in a problem-domain context where conflicts of interest were so extreme and intractable. In the end we decided to retain the paragraphs that described the model-stakeholder fusion for four reasons.

1. The idea of a ‘model-stakeholder fusion’ was clearly consistent with the call, the text of which ended with the sentence: “*The involvement of relevant stakeholders is highly recommended.*” If we made it clear that this aspect of the work was a high-risk, high-gain activity, we would be justified in devoting resources to the work and might learn something of lasting value.
2. We could increase the probability of success by not pre-judging the question of what a model-stakeholder fusion would look like. Instead, we would work to achieve the highest possible level of integration between weaker and stronger stakeholders. If successful, we would write a series of reports about what model-stakeholder fusion turned out to be; if unsuccessful, we would write an honest analysis of how we failed in the hope that future researchers would be able to learn from our mistakes and do better[9].

[8] *http://owsgip.itc.utwente.nl/projects/complex/index.php/2015-01-29-08-35-56*

[9] *If you wish to read those reports, see COMPLEX Deliverables 6.3, 6.4, 6.5, 6.7 and 6.10*

3. The COMPLEX team contained a mix of people, some of whom had substantial experience of integrative research, whilst others had almost none. We could use the internal stakeholder community reflexively as a 'test-bed' for hypotheses about how best to achieve a model-stakeholder fusion with external stakeholders.
4. We had a fall-back plan. The process of writing, negotiating and executing a successful research project would require us to write a detailed Description of Work (DOW). The DOW would, in effect, be a model of the project that reconciled the needs and perceptions of a number of relatively weak stakeholders (the individual scientists working on the team) to those of a much stronger stakeholder (DG Research). If we were able to carry this work from inception through negotiation to final delivery, we would have at least one successful model-stakeholder fusion to describe. This deliverable - D6.12 - was written in to the DOW to provide an opportunity to describe that work and the second author was recruited as a volunteer to help research it.

3.1 Complex Causality and the Three Time Perspectives

Clearly COMPLEX could not build an "über-model" to cover all aspects of the climate-energy-economic system. Instead we decided to design a flexible system of integrated models and components that could accommodate socio-natural complexity. In particular, we said we would pay special attention to situations where counterpoised pairs of cause-effect relationships could be valorised and scope for innovation might be found. We described these situations as instances of *complex causality*, a concept we will use repeatedly in the remainder of this report.

In the classical sciences causal relationships are space-time invariant, but in complex systems research, the relationship between cause and effect is linked to space-time perspective. Each perspective brings some phenomena into the foreground and backgrounds others. The relationship between cause and effect is often clear within each of these *quasi-classic*[10] domains, but may be reversed or confounded as space-time perspectives change. Complex causality is not a philosophical construct - it echoes Nils Bohr's ideas about deep truths. The only reason it seems a little odd is that most disciplinary communities tend to keep these perspectives in separate categories and to defend them from refutation as if they were trivial truths.

Consider, for example, a limestone catchment. Viewed from the deep time perspective of geology, it is clear that water movement has cut through rocks to create limestone gorges, fissures and underground rivers. Surface water redistributed sediments to create new landscape features. Water movement *causes* landscape structures. When the same landscape is viewed on the meso-scale of hydrology, however, it is equally clear that persistent landscape structures - riverbanks, underground cave systems and flood plains, say - constrain the passage of water through the catchment. The shift of perspective seems to transform cause into effect and effect into cause. It corresponds, more or less, to the distinction of geomorphology from hydrology, each of which brings one of these deep truths into the foreground and backgrounds the other.

Coming down a step further to the micro-scale, we see that the system can flip from one set of causal structures to another in response to environmental perturbations and human action. Indeed, humans often manipulate causal structures, for example by building dams and harnessing hydro-power, by clearing forest or abstracting water from the aquifer. Although unforeseen and unwelcome events occur in hydro-geological systems - flash-floods are an obvious example - they can be managed by re-shaping human activity systems in ways that mitigate or accommodate system shocks by modifying the space-time signature of the system and switching control between the poles represented by two deep truths.

Innovative hot-spots often occur in situations where causality is complex and the three space-time perspectives of *Annales* historiography can be valorised. There is a *deep-time perspective* with one set of cause / effect relations and a *synergetic conjuncture*, where cause / effect relations are usually very different and often appear to be time-symmetric. Finally there is a micro-scale where narrative chains

[10] *The physicist Murray Gell-Mann uses this term in his book The Quark and the Jaguar*

of *events* occur, some of which can re-organise the system from the bottom-up by flipping it from one conjunctural attractor to another.

In the natural sciences, which deal with situations where human agency is limited, conjunctural attractors can often be characterised experimentally and predicted *ex ante*. If the power flows, the light bulb will emit light, if not, the bulb will be dull. In human activity systems, however, the new conjunctural attractors may be contingent on ideas we have not conceived and knowledge we do not yet possess. The difference is that between metastability and innovation or, perhaps more helpfully, between uncertain and meaningless predictions.

3.2 Management, Regulation and Innovation

The COMPLEX Model Repository contains a single instance of a third type of model. It was not a computer program or a database or even a formal description of some process, but a short book titled *The Behavioural Ecology of Project-Based Science*[11]. This book contextualises the COMPLEX project by describing some of the techniques and ideas that could be used to increase the likelihood of success in integrative socio-natural science.

As mentioned earlier, the COMPLEX project and each of its workpackages are human activity systems and, as such, reflect (i.e. have the same ontological structure as) the human activity systems they study. Reflexivity can be confusing, at least until you get used to the idea of fractal networks of systems, each with its own characteristic behavioural roles and space-time signature, but it is also an advantage. The project can be used as a 'test-bed' for experiments in innovation-management and integrative research that would be ethically indefensible in other contexts.

Figure 1: The Two-headed god Janus

https://en.wikipedia.org/wiki/Janus#/media/File:Janus1.JPG

A project co-ordinator can build 3 or more of these 'test-beds' in the course of a decade and curate insights from one to another. COMPLEX, as we have already indicated, was just such a test-bed. We have used it to valorise the concept of a model-stakeholder fusion.

COMPLEX maintains a distinction between two types of administrative activity that give our project a Janus-headed structure: the inward-looking face of the administration-Janus is responsible for the project's day-to-day management. *Management* tends to be responsive, fluid and largely automatic. When a car is being driven within conventional safety standards, for example, the skilful driver is managing a continuous process that involves a complex of small-scale, semi-automatic responses.

The outward face of the administration-Janus looks towards the institutional context within which the project is located. It is responsible for negotiating the 'boundary conditions' within which the project will operate. These constraints define the basis on which the project will be *regulated*. Regulation is usually organised around auditable targets - objective measures of compliance and system health. On EU projects, for example, the list of project Milestones and Deliverables, together with the accounting and reporting conventions set out in the General Contract, and the specific terms and conditions laid out in the Consortium Agreement provide the regulatory framework for the project. These boundary conditions set limits beyond which the project may not go.

[11] *https://www.researchgate.net/publication/261175321_The_Behavioural_Ecology_of_Project-Based_Science*

The car analogy used above to describe managerial action can be extended to regulation. A car driven within boundary constraints is being managed - the ride should feel smooth as the driver and the vehicle respond flexibly to changing circumstances. However, if speed increases dangerously or the road becomes slippery, then the system shifts from a managerial to a regulatory mode as the driver makes rapid corrections to avoid driving off the road or drifting into the wrong lane. A process that depends solely on managerial action is usually safe, predictable and controllable. When regulatory boundaries are violated, however, unforeseen consequences are likely and some sort of top-down, hierarchical action is required. Sometimes these emergents are unwelcome - like the erratic movements of a car out of control. Sometimes, however, they are the source of new insights and innovations. Occasionally, both types of circumstance co-occur as an unwelcome emergent triggers domino-chains of reaction and accommodation that, in turn, change perceptions and habitual behaviours.

When the boundary conditions are violated, a warning sounds and the project must be re-oriented. Integrative projects are often high-risk, high-gain ventures and it often happens that one or more boundary conditions cannot be satisfied. When this occurs, an explicit process of re-negotiation is initiated that sets new boundary conditions. Key external stakeholders must be involved in this process. In practice, DG Research maintains a distinction between modifications that can be signed off by the Project Officer and those which require that the contract be re-negotiated. This effectively sets boundary conditions on the boundary conditions and amounts to a moratorium on some types of adjustment. In an institutional context the regulator would be called something grand like: 'Board of Directors' or 'Senate', but integrative projects are transient, heterarchical consortia, and regulators usually consist of a steering group with representatives of key external stakeholders.

One of the most challenging tasks for a project regulator is that of distinguishing a research activity that has gone off course as a result of bad luck or bad management from one that stands on the threshold of a valuable innovation. The validation problem is significant here. As explained earlier, each project is regulated and managed in respect of a discursive model called the Description of Work (DOW). When it was first written, the COMPLEX DOW was an aspirational document - a model of a project-system that did not yet exist, but which we would like to bring into being. The DOW was a meld of *ex post* experience of project-based science and *ex ante* aspiration. Those aspirations included the desire to innovate.

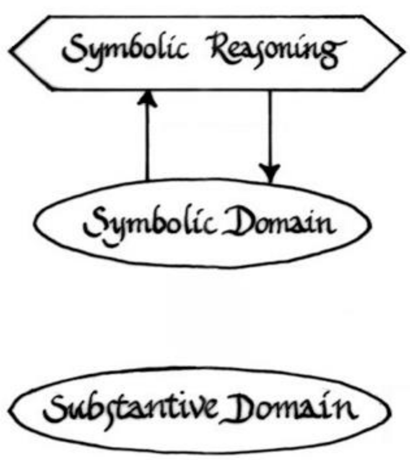
Once the proposal had been evaluated, selected for funding and the project negotiated into existence, the DOW became, in effect, a *decision-support model* that would be used as a basis for managerial and regulatory action. This shift from *ex ante* aspirational model to *ex post* decision-support system requires a little flexibility. If we were to treat the DOW as a well-validated blueprint of what the COMPLEX project should be, we would run the risk of over-regulating the project and stifling innovation. If, however, we were to ignore evidence of unplanned mission-creep, the project would be rightly judged a failure. A balance must be found.

For inexperienced co-ordinators these decisions are often one-shot operations that either go well or go badly. As experience is gained, however, the co-ordinator develops insights into the research process that can enhance the probability of a happy outcome. Some co-ordinators try to avoid conflict at all costs, developing fuzzy measures of success or failure and working round non-delivery or system failure in the hope that everything will turn out alright. Others favour sharply bounded, objective measures and enforce every detail. The present writer's experience is that conflict-free projects seldom deliver high-impact results, but that excessive conflict creates discursive 'noisiness' that undermines trust. In practice, it is helpful to develop general, theoretical models of innovation-processes based on experience gained over many similar projects, as we have done with COMPLEX.

4. Modelling the Modelling Process

This section will develop a generic model of the modelling process, a task that would be impossible were it not for the reflexivity of integrative systems. These models will take the form of annotated diagrams. A wide range of diagramming conventions is used in systems analysis and it may be worthwhile to observe that these diagrams are to be interpreted cybernetically. The arrows represent information flows[12].

Figure 2: a dynamical system used as a model of some substantive domain.



This diagramming convention allows us to provide a simplified representation of a model that can be applied to computational and qualitative models without loss of generality. A *model* is a symbolic representation of some substantive research domain. The act of modelling requires us to develop some sort of symbolic notation that allows us to represent the state of the substantive domain as a collection of symbols. These symbols can be words, numbers, diagrams or switch-states in a computer.

Every systems model establishes an analogy between two domains that correspond, more or less, to the distinction of theory from hypothesis. The *substantive domain* represents our knowledge or beliefs about a particular arena of activity. It is usually represented by a description of system ontology, including an informal inventory of things and categories of things and a state-description at some time, *t,* supplemented with additional information about the dynamic processes that transform the system's state through time. Our knowledge about the substantive domain tends to be provisional, and modellers are psychologically well-prepared to revise and update these hypotheses.

The theoretical component of a model appeals to knowledge about a complete genus of systems of which the study arena in which we are currently working is an instance. Theoretical knowledge can be thought of as a s*ymbolic domain* that allows us to describe the system as an instance of a general class with its own characteristic ontology and current system state.

Modellers often appeal to our intuitive understanding of physical spaces and use it to characterise some abstract and occasionally difficult ideas. Just as a geo-space is a set of possible locations in a landscape, so a 'state space' is a set of possible states. This spatial analogy allows them to harness mathematical ideas about dynamical systems and use them to build formal models.

In mathematics a *dynamical system* is a mapping of some 'space' onto itself. You can think of that mapping as a symbolic manipulator that receives information describing the system's state (location) at time $t=n$ and delivers information that describes its state (location) at $t=n+1$. By connecting the output of the information-processer to the input, you can iterate and re-iterate the program and predict the system's state at $t= 0, 1, 2, 3, \dots$ and so on.

Dynamical systems can be used to simulate a time-series or trajectory of movements through the 'state space' of some system. Under certain circumstances we can use our spatial intuitions to develop a rich language that describes types of location in a state-space. An *attractor*, as mentioned earlier, is a

[12] *The concept of information used here is not that used in Shannon and Weaver style information theory, which measures information in terms of binary integers. As C H Waddington pointed out in his book Tools for Thought, the telegraph messages:*

MEET HIGH MARKET TWELVE TEN and MEAT HIGH MARKET TWELVE TON

only differ in two characters and therefore have similar Shannon / Weaver information contents. When one places them in context, however, one is an appointment to meet at the junction of High Street and Market Street and the other an instruction to offload 12 tons of rotten meat.

In human activity systems, information is a data-stream that changes the recipient's state. A data stream that has no impact is not information. In practice this means that the data flow itself is often less significant than the interpretive context or framework the recipient uses to interpret those data.

region of the state-space that tends to trap passing trajectories, while a *repellor* is an unstable region that tends to drive trajectories away. A *basin of attraction* is a region that funnels trajectories into an attractor. We can even speak about different types of attractor - point attractors, say, or cyclical attractors.

Although attractors, repellors and basins are easy to visualise in simple geometrical spaces where dynamic trajectories are smooth, continuous and well-behaved, modellers often use these terms in more complex situations where it may be impossible to operationalise geometric ideas like distance or curve. The language of dynamical systems survives the transition to these non-metric spaces, though our spatial intuitions may be challenged and pop-science books often speak of 'strange attractors'. The use of computers is very important when state-spaces get messy and dynamic processes can no longer be reduced to smooth, continuous curves. Although quite simple dynamical systems may be analytically intractable, it is often possible to use computers to characterise attractors, repellors and basins in complex systems.

The iterative methods we have described for dynamical systems can only be guaranteed to work if the state-space is 'closed' under transformation, i.e. if the product of every symbolic manipulation is a location in the same state-space. If you were to input the state of the economy today and the output named the winner in a horse-race last week, then the model would be very difficult to operationalise, even with the help of a computer. Classical systems modelling method requires that the symbolic manipulator *maps the space onto itself.* This constraint can be thought of as a *boundary condition* that restricts our attention to dynamical systems whose state-spaces are *logically closed* or *bounded* under symbolic manipulation.

However the substantive domains that engage the COMPLEX project are not logically closed - they are capable of innovating. In an innovative system, the state-space is not bounded under symbolic manipulation. The key idea is that our models (symbolic domain with symbolic manipulator) influence our actions in the substantive domain. Those actions may have unforeseen consequences that force us to revise our models. In this way, we establish a reciprocal flow of information between the substantive and symbolic domains that allow them to co-evolve. Instead of speaking in terms of state-spaces, this approach requires us to contemplate 'model spaces' in which every location corresponds to a different model and there are reciprocal information-flows between the substantive and symbolic domains.

4.1 Example 1: Model-Stakeholder Fusion

Stafford Beer[13] observed that we humans do not solve everyday problems by trial and error, for example we do not say to our children 'Quick, run across the road. Damn! They died.' Rather we run mental simulations designed to help us anticipate and manage risks or opportunities and teach the children how to behave. The process of developing a model-stakeholder fusion can be envisioned in these terms. Our aim is to negotiate a consensus about dynamic processes, risks and opportunities. Every model-stakeholder fusion is the product of an *ex ante* (i.e. forward-looking) mental simulation that represents stakeholder understanding, aspirations and concerns.

A model-stakeholder fusion usually relates to a reflexive problem domain or *arena* that contains, amongst other things, a fractal network of stakeholders, some of which are actively involved in the modelling process. All stakeholders (modellers or not) have, at their disposal, an unbounded 'possibility space' of behaviours and responses that represent their *operational* competence. The model receives information from the arena, processes that information internally and transmits new information to its stakeholders that modify their operational responses. These responses change the state of the arena, which may transmit further information to the model, triggering another cycle of information-flows.

As explained earlier, models have two domains: a *substantive domain* that represents the stakeholder's current hypotheses about the arena at hand and a *symbolic domain* that locates the arena in a more general theoretical context. In many of the COMPLEX project's case-studies, the process of symbolic

[13] *(1979) The Heart of Enterprise*

reasoning is simulated by a computer program and the state-space the model maps onto itself is *logically closed*, or *bounded* under transformation. However, the model-stakeholder fusion we describe here represents an arena that contains the modellers, together with other stakeholders, with each stakeholder community forming part of a fractal network that incorporates large parts of the system being modelled.

As stakeholders work on the model, the negotiation process can change their understanding of the arena in ways force them to revise the substantive domain of the model and even define qualitatively new state-spaces. This logical openness implies that some, at least, of the symbolic reasoning must be carried out by stakeholders outside the model itself. Some of our case studies, for example, used participatory modelling techniques in which stakeholders were actually invited to change the model's ontology or suggest different types of cause-effect relation. They might be invited to re-draw maps so as to simulate changes in planning regulations or the effects of installing renewable energy generators on their own homes, for example or to perturb the system's state by imposing a carbon tax. In this way stakeholders engaged in a process of negotiation, both among themselves and with the model, that re-shaped both the model and stakeholder mindsets in real time.

Similarly, the process of writing the DOW involved a lengthy program of negotiation with participating scientists. Text would be drafted, evaluated, rejected or re-drafted until the draft seemed plausible, operationally practicable and scientifically worthwhile. We were not aiming for an historically accurate description of the COMPLEX project. Rather, we were running mental simulations, *ex ante*, that would characterise future threats and opportunities. The finished DOW was a model-stakeholder fusion that would form the basis for co-operative action at the next stage of our work.

Figure 3 (overleaf) is an attempt to realise this complex, reflexive structure in the form of a cybernetic diagram. Whereas, in Figure 2, the model is represented by a symbolic domain coupled to a symbolic manipulator, the situation in Figure 3 is rather messier. We have used a dotted line to approximate the model boundary, which now consists of a symbolic domain linked to a substantive domain by reciprocal information flows.

Here, as in Figure 2, the *substantive domain* that describes the ontology of the arena together with our best understanding of system dynamics, threats and opportunities. There is also a more or less fixed theoretical framework, the *symbolic domain*. Most computer models have an internal symbolic manipulator, but some, at least of the symbolic manipulation, now takes place outside the model. The current state of the model is transmitted as information to the modelling team - labelled 'Operation' in this diagram - which works with other stakeholders in the model's Arena to interpret and respond to that information.

Whatever else it may be, a model-stakeholder fusion is much more complex than the mapping of a bounded state-space onto itself. The model represents an arena that contains the modellers and the model itself. As the model is revised and developed, it changes mindsets among some stakeholders which, in turn, changes their behaviour and modifies system behaviour. This is a perfect example of what we have been calling complex causality. Like the landscape which is created by water movement and also constrains the movement of water, so the modellers use their understanding of system structure to create a model which changes their understanding, which changes the structure of the system being modelled.

The 'model-space' explored by this model-stakeholder fusion is logically open and unbounded. That word 'unbounded' may give a false impression. When we use it, we do not mean that absolutely any conceptual model, even irrational or impossible models, can be developed this way. In practice most of the modifications made to the model revised hypotheses about the substantive domain and modified the model's internal symbolic manipulator. The symbolic domain that represents scientific theory tends to be much stabler than the substantive domain. This means that the 'model space' is *unbounded* in the sense that stakeholders can use it to innovate, exploring ideas about model ontology and dynamics they would not have been able to represent at the outset, but it is not *boundless*. There is a backstop of theoretical constraint that contextualises the work. This theoretical backstop is an institutional constraint that limits the project's agency.

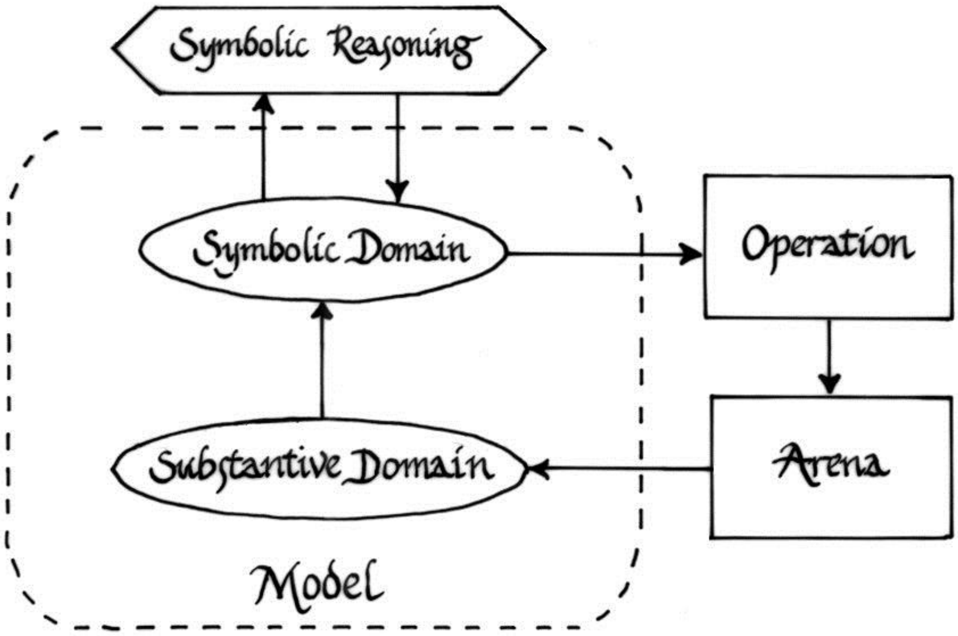


Figure 3: a Model-Stakeholder Fusion in which operational decisions are facilitated by gathering information about the state of some system (substantive domain) translating that information into symbolic form (usually as data) and effecting some analysis.

4.2 Example 2: Decision Support

Although this report deals with modelling in general, the model that has most strongly influenced it is the DOW, written 4 years ago to describe the work the COMPLEX project would undertake, its Milestones, Deliverables and expected impacts. There are three reasons for this emphasis. Firstly, one of the writers is very familiar with the DOW, having spent in excess of 2,000 hours negotiating the model-stakeholder fusion and co-ordinating the program of research. Secondly, both writers have devoted a considerable block of time to the task of dissecting the theoretical domain of the DOW out and presenting it separately. The DOW, like all models, is a hybrid structure, with hypothetical and theoretical domains, but the accompanying monograph on the Behavioural Ecology of Project-Based Science is almost pure theory. Consequently, the DOW is much messier and more conjectural than the monograph, which has been extensively validated and fine-tuned to past experience.

The monograph was the product of more than 25 person-years research on the applied anthropology of integrative socio-natural science. Our ideas have certainly evolved over those years, but the rate of change has definitely slowed. The principal author, for example, learned most of what he knows about project design and co-ordination more than a decade ago. Over the last decade he has learned that cybernetic diagrams are very useful ways of communicating ideas about project management, that projects managed to minimise internal conflict seldom innovate, and that high conflict levels usually destroy trust to the point where innovation is impossible. Small to medium projects (250,000 - 2,000,000 €) are much easier to keep within these bounds than large ones, in part because work-packages become self-sufficient and researchers have fewer incentives to look outside and in part because the combinatorial structure of large projects makes it much harder to service bilateral and trilateral relationships. The projects on which he worked over that decade were productive and the theoretical lessons learned were valuable, but theoretical framework was established early on and has proven stable.

The third, and most important reason for focussing on the COMPLEX DOW is that it has passed through two distinct phases of existence. In the first phase, the DOW was used *ex ante* to develop model-stakeholder fusion as described in 4.1 above. In that phase, the COMPLEX project only existed as a collection of ideas, aspirations and hopes. It took about 12 months to take the DOW from inception to negotiation. With 7 workpackages and 17 institutional partners, the DOW is a substantial document (113 close-written pages) and the process of researching it and steering it through negotiation was arduous and time-consuming. Since the EUs evaluators awarded the proposal a very high score and all the partners felt the DOW was an accurate and workable description of the system they wished to create and manage, it seems reasonable to describe the finished DOW as a successful exercise in model-stakeholder fusion.

In the second phase of the work, we switched from an *ex ante* to the *ex post* modus operandi by finalising the DOW and incorporating it into a contract with DG Research, strengthened with a legally binding Consortium Agreement. Although the DOW contained a full list of the Milestones and Deliverables that comprised the project's boundary conditions, there could no suggestion that the model had been validated. None of us had worked on the COMPLEX project before, because it hadn't existed. All of us were hoping that the project would innovate by finding new ways of thinking about the transition to a low-carbon economy and new strategies for facilitating that transition. This hope created the usual challenge for the project's steering board. How were we to distinguish between a project that deviated from the DOW through ill-luck, poor management or mission-creep from one which deviated from the DOW because it was capitalising on an emergent opportunity for innovation?

The strategy used within COMPLEX was to mark the transition from *ex ante* model-stakeholder fusion to *ex post* decision support by imposing boundary conditions on the DOW. The boundary conditions, as we have explained, were pre-agreed, auditable indicators of compliance and system health. When those boundary conditions were subsequently violated, the hypothesis represented by the model was taken to have been refuted, and the project's knowledge-state changed. The model had to be revised to accommodate the new knowledge. Usually the effort of revision was trivial - mistakes had been made or plans had failed and regulatory action was taken to nudge the project back on-course. Occasionally, however, an opportunity was found to do things better and resources were shifted to reinforce the deviation.

The resulting model, a *decision-support tool*, included all the elements of the model-stakeholder fusion just described and an additional component that describes and formalises the positive / negative feedback switch-state that was activated whenever boundary conditions were violated. See Figure 4 overleaf.

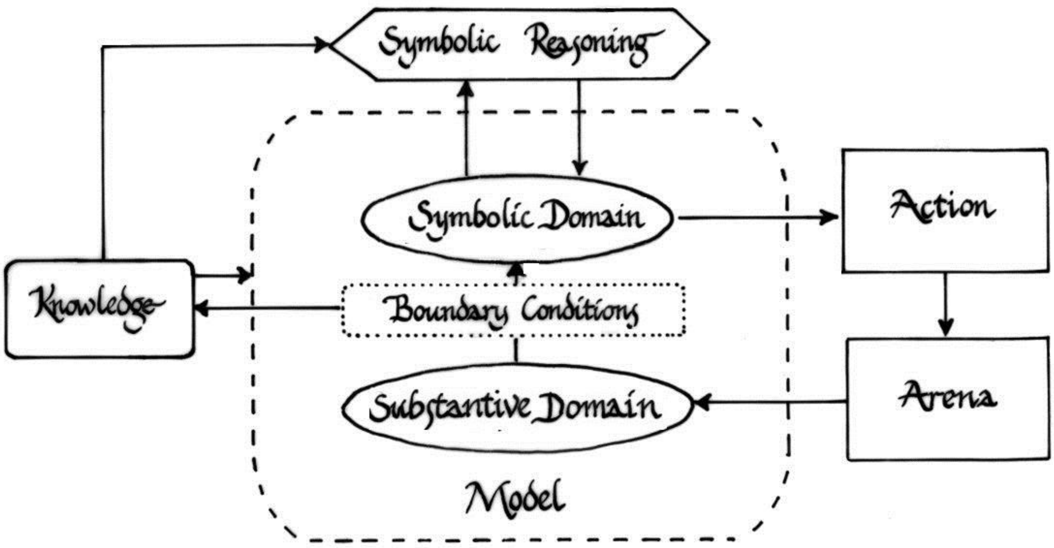


Figure 4: A stable decision-support model with boundary conditions

4.3 Distinguishing Mission-Creep from Possible Innovation

Regulators need to devote a little time and effort to distinguishing between careless, unlucky or incompetent system management and innovation potential. In the latter case, the violated boundary condition corresponds both to a loss of control and to a perceived opportunity to do things differently. The recipients of these insights usually experience a strong, affirmative feedback stimulus that simultaneously changes their understanding of how the system works and what their role in that system should be. Although epiphanies are often experienced as sudden insights, they invariably follow an extended period of research effort and contemplation. That research process probably involves the formation of new patterns of neuro-connection in the brain. The *Aha!* moment occurs when some event activates a cognitive trigger that brings those new circuits on-line.

The regulator can usually distinguish system-failure from innovation because those responsible for managerial control become excited and enthusiastic. They start talking about opportunities they had never imagined and completely new ways of doing things. Sometimes they actually experience an identity-change comparable to a religious conversion and dedicate part of their lives to realising this opportunity. Above all, they take possession of the new ideas and demand the opportunity to strike out on a new track.

When projects run out of control, in contrast, researchers are more likely to cover their tracks, shift blame and responsibility, or pretend that everything is going to plan - they haven't really gone off course - "Yes, we know that the research schedule says we should be doing this, but what we have done is at least as good". If the project has run so badly off course that everyone can see deliverables slipping, they may start trying to negotiate more resources - they need a little more time, more data, a bigger budget, another chance, ... next time things will be different.

In most circumstances where boundary conditions are violated, regulators will select the negative feedback mechanism that forces the project back on-course, either by clawing resources back from the failing team or by taking over responsibility for task management. This negative feed-back, whether it be a private instruction to 'straighten up and fly right' or a formal Compliance Notice issued under the Consortium Agreement, is almost always perceived as some sort of punishment. There will be grumbling, a loss of confidence and trust and, under extreme circumstances, complaints about the unreasonableness of the project co-ordinator and project officer.

This negative feedback mechanism always makes the project harder to regulate and often has unforeseen and unwelcome consequences. It helps if the Consortium Agreement has real teeth because uncooperative individuals will usually be over-ruled by their line managers within the partner institution. Sometimes, alas, the individuals involved refuse to co-operate altogether and the regulator must decide whether to tighten the screws further or write some part of the shared work off. There is no decision harder than the decision to write off sunk capital and move on.

Of more interest here, however, are those rare occasions when a research team encounters an opportunity to innovate and the regulator needs a backup system - a positive feedback mechanism that reinforces the deviation and allows the team to explore a new conceptual model. The wonderful *Aha! moment* that accompanies the acquisition of new knowledge provides a sort of litmus test that regulators can use to distinguish a potential innovation from a probable failure. With a potential innovation, those involved become enthusiastic and 'take possession' of the deviation, arguing persuasively that it should be reinforced and rewarded.

5. A Concrete Example

The first four sections of this report summarise the conceptual and theoretical framework developed for innovation-management in COMPLEX. This section describes a concrete example of the way the method was deployed in practice.

Writing the COMPLEX DOW was an exercise we, as a consortium, did together and one might imagine that all those involved in this process would have the document well fixed in memory.

However, the experience of this project and others has been that this is not the case. The DOW is often perceived as a hoop-jumping exercise - once we have the contract and the grant, we can just do our research and everything will all fall neatly into place. When the project runs off course, those involved tend to take refuge in their research interests and habits, re-framing their memories of the DOW to accommodate this drift.

Scientists with little experience of international, competitively tendered research are particularly susceptible to this forgetting. They understand intuitively that the DOW must promise high-impact innovative research, but when the pressure is on to deliver and deadlines are slipping, retreat to laager in a way that actually reduces the likelihood of a successful outcome. The more one tightens the regulatory screws, the less safe they feel and the less likely they are to deliver innovative, high-impact results.

It is as though research projects have two deep dynamic attractors; one offers the scientists a great deal of managerial freedom by pushing responsibility and accountability as far down the delegation-chain as possible. The idea is that if people feel safe and free to experiment, new ideas will bubble up from the research and some of these will create new opportunities for future action. This is the *heterarchic* attractor, characterised by distributed processing. The second attractor is heavily regulated and *hierarchic*, with low trust-levels. It is more angst-ridden, conventional and defensive.

It would be easy to equate heterarchy with innovation and hierarchy with paralysis, but actually both causal mechanisms must be present in all cases. Without strong regulatory action, the team has too much managerial wriggle-room; the project will drift off course and quality standards will be low. What seems to happen is that some unwelcome event (a violated boundary condition) triggers a regulatory response that bounces part of the project from a relaxed, disciplined heterarchy into a more hierarchical, defensive attractor and the consortium needs to find some way of salvaging the situation.

By the end of Period 1 (18 months into the project) boundary alarms had rung twice and internal quality regulation procedures had been used to re-establish control and get the project back on-course. Fortunately these problems could be contained within a particular task-group or WP and corrective action was duly taken. In practice, regulatory protocols have a strong, hierarchical structure as the co-ordinator, with the backing of Steering Board members and external funders, puts a task group or WP into 'special measures' designed to bring it back on course. As one might expect, these special measures were perceived as punitive; the more experienced internal stakeholders got over the discomfort fairly fast, but less experienced personnel clearly found the experience distressing. Resources were mobilised to bring the project back on track and we entered Period 2 in good order.

By the end of Period 2 (3 years in), however, DG Research rejected our scientific report because of unexplained delays to some deliverables. Some of these deliverables were strategically significant because they were related to the process of integrating the research effort. The triggering event had been identified by external quality audit, so immediate action was required. The process of redrafting the scientific report became a priority and arrangements were made for a meeting between the Project Officer and co-ordinator. Although one has to consider the possibility that a boundary alarm heralds a valuable innovation, initial indicators were discouraging. There was no evidence among any of our personnel of a glorious epiphany or of persuasive demands that the deviation from DOW be reinforced, just a sorrowful acknowledgement that this part of our work together had gone off-course. Nobody felt a burning need to take possession of these deviations or change the way we worked. Rather the focus was on emphasising our (very real) achievements elsewhere and making the best of the rest.

Regulatory constraints always move a project closer to the hierarchical attractor, with instructions handed down the chain from funders to co-ordinator and downwards to WP- and task leaders. Task- and WP leaders become personally accountable for non-delivery and their employers may actually be required to return grants or transfer resources to other partners to make the shortfall good. It isn't in anyone's interests to allow the situation to escalate this way and a decision was made to use our fourth and final plenary meeting to see whether the project could be bounced away from that hierarchic attractor towards a more relaxed, heterarchical configuration in which people might possibly experience epiphanies and find innovative ways of carrying the work forward.

It is perhaps worth emphasising the seeming unreasonableness of this plan. The project was unlikely to fail - we had achieved too much - but these few integrative tasks had drifted so far off course that nobody felt able to argue that all was well or that a quick fix was possible. We had almost enough deliverables 'in the bag', as it were, to carry the project to completion and so could have elected to use the regulatory disciplines described in the Consortium Agreement to force partners into compliance or claw resources back. COMPLEX had been a very ambitious project and we had achieved a great deal. DG Research knows what the phrase 'high risk, high gain' means. We could have negotiated a compromise on these last few deliverables.

Instead of taking that conventional route, we decided to stage an experimental action designed to bounce the project into a happier, more heterarchical modus operandi. These sorts of actions are always a long-shot because you aren't trying to steer the team towards a well-defined and anticipated goal. Instead, you are trying to create circumstances where epiphanies - those utterly surprising and exciting *Aha! moments* - are more likely to occur. At a point in the project cycle where commonsense and common practice both require you to tighten the regulatory screws and kick some bottom, we decided to create a safe environment for all our colleagues to play in and deliberately lose control of the project. This is not an easy call to make, partly because the likelihood of failure is rather high and the consequences of failure are potentially embarrassing.

Normally with reflexive experiments that use the research team as a test-bed for new managerial strategies, one seeks informed consent from partners, but in this case we had to be satisfied with uninformed consent. To tell colleagues that the purpose of a meeting was to loosen hierarchical control and see whether they could come up with a game-changing innovation would create a level of performance anxiety that might stifle innovation. So the team was invited to consider using the 2-day meeting in Sweden to prepare an open stakeholder engagement event and individual volunteers were sought to facilitate that work. There was no shortage of volunteers and most of the COMPLEX team agreed to attend the open day to welcome and work with external stakeholders.

5.1 The Problem

In addition to its regionally focussed case-studies, COMPLEX was to have specialist workpackages for model development (WP5) integration (WP6) and Dissemination (WP7). The integrative workpackage would be responsible for building dynamic couplings between model components, would develop methods for comparing the output of different models of the same system *and* would develop the model-stakeholder fusion. It was the last of these tasks, the 'model-stakeholder fusion', that had gone off track. WP6 was supposed to delegate staff to attend stakeholder meetings and help fine-tune modelling efforts, but this hadn't happened. Because of this, information-flows between our regionally-focussed WPs (2, 3 and 4) and WP6 had been weak and the two software packages we had promised to develop that would facilitate model-stakeholder fusion in the future had not been developed.

The results of our work with external stakeholders are described elsewhere (see D 6.4 and Volume I of the Scientific Report, *in prep.*). COMPLEX achieved a very high standard of integration with external stakeholders, but the managerial framework for internal integration must be judged a qualified failure because WP6 personnel did not get to enough of those meetings. Our problems seem to have stemmed from the following factors:

- The regional case-study teams established links with external stakeholders early on. In some cases they had had long-standing relations with external stakeholders and simply called them up. WPs that were well-equipped to combine formal models with stakeholder engagement already had the modelling competence they needed in-house. It was easier for their own modellers to work with their own stakeholders in their own languages than to call on the resources of a specialist, integrative WP6 because there would be language barriers and problems of communication to overcome.
- Not all of our regional case study teams were well placed to co-operate with WP6 on these tasks. WP3 was, in many cases, the easiest case-study to integrate because it had experience of work with very diverse external stakeholder communities and personnel had been using participatory modelling approaches for some years. WP2 had strong pre-existing links with

technocratic stakeholders and was well used to developing decision-support models, it didn't need support from WP6.

- WP4 was the least experienced of our case-study teams, with almost no experience of participatory- or decision-support modelling. Although it had long-standing links with some external stakeholders, WP4 had had little experience of an integrative approach at project start-up, and there were differences in the conceptual understanding of the process. There were also recruitment problems which delayed some WP4 tasks. At the first stakeholder workshop held in Period 1, colleagues from other WPs with experience in integrative method participated and offered guidance and advice. It appeared that WP4 was not yet ready to follow through in all parts, but eventually experienced personnel were recruited locally and some excellent engagement work was done. WP4 personnel who originally were not quite ready, later played pivotal roles in our stakeholder engagement work, but there were delays and receptivity barriers to overcome.
- On a number of occasions the provisions of the DOW were forgotten. All our regional case-study teams had time-budgets divided between their own, regional WPs and WP6, but they tended to forget or fail to realise that all their model-integration work belonged to WP6. This is a common problem in Framework projects and will probably continue to afflict H2020. Evaluators and accountants expect consortia to put management and research activities in different WPs, and to have special, over-arching WPs for tasks like integration and dissemination, but this is actually a very unnatural and counter-intuitive way of organising the work. In reflexive socio-natural science, the project team can be used as a test-bed for research methods and the distinction between research management and research method simply disappears. Scientists working on a regional case-study find it hard to 'switch hats' when they move from research activities to integration because there is no sharp boundary or warning sign:

  "DANGER! EXTERNAL STAKEHOLDERS APPROACHING!"

  As a result of this, activities that could have been assigned to WP6 *de facto* were not reported through WP6 and that made WP6 look weaker than it was. Although WP6 task leaders repeatedly asked to be involved in work with external stakeholders and these requests were reinforced by the co-ordinator, the response from case-study teams was disappointing. Connections were made between WP6 and WP3 fairly early on, but other links were weak; some WP6 deadlines had slipped and boundary conditions were violated. Our second scientific report was rejected on the grounds that these delays had not been adequately explained.

5.2 What were we hoping to achieve?

The purpose of our experiment was to create an environment that was propitious to innovation. Innovations take a long time - it takes time to grow new networks of neuro-connection and time to capitalise on new ways of thinking and understanding, but we had an advantage. Our research team had been working diligently for three years and so had probably developed a few new connection-patterns. If we could join forces with external stakeholders who were themselves interested in in the low carbon economy and create an environment that was safe enough to allow people to take small risks and stimulating enough to trigger epiphanies, we might possibly get some interesting results.

We designed our open day of stakeholder engagement around a set of managerial principles described in our book as follows:

> *It is a good idea to keep people working in teams of like-minded specialists most of the time, so their shared knowledge-base is robust and unchallenged; bring them together in small groups; create a safe, no-blame culture, exclude incorrigible 'networkers' and high-profile spectators – they make the team noisier and harder to manage. Provide opportunities for social interaction with good food and make sure there is space for break-out sessions. Avoid hedonistic pleasures like poison.*

> *Focus the group on one or more specific, auditable deliverables. Don't bring them together to discuss the problems of child poverty or global warming, or to decide whose fault it is that public investment in science has unforeseen and unwelcome consequences. Concentrate on some small itch that everyone wants to scratch and link it to a concrete deliverable.*
>
> *… Start with social interaction in small groups and negotiate a shared task for the whole group to tackle. When people come together, you need them to leave their cultural hang-ups behind and focus on that shared target. Once that target has been achieved, or substantial progress has been made, you encourage delegates to take possession of part of the shared task and carry it back into their own cultural home-ranges. Just about the silliest thing you can do is to create a trans-disciplinary circus and expect people to operate outside those comfort zones indefinitely. They will fight. There will be winners and losers.*

The COMPLEX team would be given 2 days in closed session and would work on three tasks:

1. To come up with a list of external stakeholders and make a plan for managing project legacy
2. To give the modellers on WPs 2, 3, 4 and 5 an opportunity to brief WP6 about work with external stakeholders
3. To plan an open day of engagement with external stakeholders

Since the strategic aim was to bounce the team from a hierarchic, discipline-based *modus operandi* into a more heterarchic configuration, it was decided that the open day would start with a brief orientation session and then split into parallel sessions, each organised by a different team. We would have four small rooms in a facility called *Refugium*, where teams would have to compete for the attention of our visitors. The rooms were small and sufficiently close to each other to allow people to see and hear what was going on in other sessions and come and go as they chose. At the end of the open day, we would invite visitors who had lasted the course to meet with us and tell us what they thought we should do next.

Some of our partners expressed frank dismay at this format, asserting that nobody would turn up to listen to foreigners talking about European research and complaining that the program was too risky. They were told that this was an experiment, that responsibility for any failure would lie solely with the project co-ordinator, and that if no more than one or two people turned up, we would go ahead anyway. If necessary, we would invite a few stakeholders along as participant observers and we, the COMPLEX team, would stand in as proxy stakeholders at each other's dissemination activities. Once this was explained, the people most critical of the format threw their weight behind the experiment, actively recruiting delegates to attend the meeting and arranging for a film-crew to be on hand to record the results. The Sigtuna Foundation created publicity material and plans were duly made.

5.3 How would we know we had succeeded?

To judge the success or failure of an experiment like this, one needs some explicit goals. Here are some candidates, starting with the least ambitious and working up.

1. To give WP6 personnel the opportunity to learn more about case-study WPs and some idea what it was like to participate with external stakeholders
2. To get enough external stakeholders to turn up to make the experience worthwhile and to invite those external stakeholders to join the COMPLEX project as advisors
3. To temporarily suspend the hierarchical structure of the COMPLEX project and see whether it can self-organise into a more organic, heterarchical *modus operandi*
4. To provide all our stakeholders, internal and external with the opportunity to experience epiphanies and *Aha! moments* of their own.
5. To focus the attention of everyone present on the needs of WP6 (Integration) in the hope of clearing the log-jam of delayed deliverables

5.4 How did it go?

The simplest way to answer that question would be to watch the film we made of the day's events:

https://www.youtube.com/watch?v=c1mfbM31fsA&feature=youtu.be

In terms of criteria 1 and 2, there is no doubt the day was a success. WP6 personnel had some valuable opportunities to see how work with external stakeholders was conducted by the various WPs and attendance was excellent, more than 20 external stakeholders joined us for the day, including stakeholders from other EU countries and outside the EU.

The first hour or so of the day was broadly hierarchical in structure, but the break-out sessions successfully disrupted that hierarchy, with each session developing a momentum of its own. The co-ordinator was particularly pleased to receive complaints from scientists that they had insufficient time to visit the other rooms. Shortly after this complaint was received, the meetings became much more fluid as scientists and external stakeholders began moving through the *Refugium*, talking, arguing and discussing other people's work. For one day, at least, the COMPLEX project assumed a more heterarchic configuration. People seemed to be engaged and having fun.

We had some evidence of success with criterion 4 too. The close-down discussion (between minutes 34 and 44 on the film) provides ample evidence of external stakeholders taking possession of new ideas and dedicating themselves to carrying them forward. We must emphasise, of course, that these epiphanies were not the direct product of work with COMPLEX, these external stakeholders have been thinking about carbon emissions over a long period and have developed new patterns of neuro-connection that prepared them well for our meeting, but the meeting seems to have provided the stimulus that brought some new circuitry on-line.

There is also some circumstantial evidence of changed mindsets within the COMPLEX team, as people have begun to realise that they have made substantial commitments to the development of model-stakeholder fusions of their own and are submitting text, evidence and reports to WP6. In the weeks that followed the stakeholder meeting, partners went through the laborious process of briefing Twente, the lead partner on this task, about their work with external stakeholders. Individual partners carried out an informal audit of their own modelling work and realised that more progress had been made on the development of those software packages than we had imagined. Modellers from OCT, for example, had encountered limitations in the commercial arena and had developed a new software tool, APoLUS, that was fully documented and ready to use. The team at IIASA had been working on the integration of model outputs and had developed not one, but two software packages. The process of documenting these was fast-tracked, giving us not two, but three software packages to deliver. It must be acknowledged that none of these packages gives the seamless, over-the web integration that we originally promised, but all three have been trialled and found useful tools for achieving model-stakeholder fusion. Research on WP3 at Twente, while it did not produce new software, had identified a combination of software and hardware that perfectly matched its needs in stakeholder engagement and WP2, which had achieved game-changing advances in managing intermittency problems in Climate-Related Energy sources, initiated a dialogue with Richard Hewitt of OCT that produced a unique record of the process of developing model-stakeholder fusion with energy companies.

We consider this experiment to have been a success and invite our peers and evaluators to make their own assessment on the basis of the film that records our open stakeholder engagement and the quality of the WP6 deliverables. However we should observe that this success could only be described meaningfully in a situation where failure is acknowledged openly.

COMPLEX is not the only project in the history of DG Research to have had an interim report knocked back for revision. Normally project co-ordinators write their final reports in a way that draws a discreet veil over missteps and pratfalls. Scientists who would not hesitate to report the negative results obtained in a laboratory experiment often find themselves in an uncomfortable position when managerial efforts fail to deliver the goods. It takes courage to take possession of these failures, to stay with the work long enough to learn something useful from them, and then to report the results of those reflexive experiments frankly.

6. Conclusion; Anticipatory Modelling and innovation management

In conventional modelling theory, model-validation is carried out *ex post*, using historical data to align a decision-support model to the evidence. Conventional decision-support modelling has a strong, *ex ante* perspective as the validated model is used to predict the future behaviour of the system under a given policy scenario. COMPLEX has built and deployed a number of these conventional models. However, we have also undertaken some participatory modelling exercises designed to negotiate a shared understanding of system dynamics, opportunities and threats. Most of these participatory exercises developed simulation models, but this report has focussed on a textual model, the DOW, which was designed to provide good regulatory control whilst leaving enough managerial 'wriggle-room' to facilitate innovation.

This participatory approach suggested a slightly unorthodox approach to modelling theory in which every model has a stable, conventionally validated theoretical domain and a fluid, aspirational hypothetical domain. Participatory modelling has a clear, *ex ante* focus, because its purpose is to negotiate a shared understanding of risks and benefits and system processes that will form the basis for future action. When the COMPLEX project shifted from participatory- to decision-support modelling, however, the focus moved from an *ex ante* to an *ex post* perspective as we continually monitored goodness of fit between the model (DOW) and the reality that was our project. This is an exact reversal of the pattern described for conventional modelling.

The idea that decision-support models are used *ex post* suggests that the term 'validation' must have a slightly different meaning in innovation-management. A decision-support model has been validated if the planned co-operative action went according to plan, and boundary conditions (measures of compliance and system health) were respected. The call to which we responded (see Introduction above) was based on the assumption that the most useful models were those that have been validated, but the analysis of innovation management presented here strongly suggests that invalidated decision-support models may actually have more innovation-potential. Every time boundary conditions are violated, the regulator needs to decide whether the project has simply drifted off-course or has found an opportunity to innovate.

Human activity systems with explicit boundary conditions are actually rather rare outside the sciences, but they do occur. The Lisbon process, for example, was a political initiative intended to facilitate innovation by encouraging European nation states to increase investment in RTD to a specific target (3% GDP). When receptivity barriers prevented them approaching this target, the Lisbon process was 'restructured' to allow the process to continue despite its manifest inability to hit targets.

Scientific research projects sometimes 'restructure' in a similar way, fudging Milestones and Deliverables in ways that effectively immunise them against embarrassing failure. The later FPs and H2020 have adopted strong regulatory protocols that discourage deceptive restructuring efforts. These regulatory protocols are an integral component of DG Research's quality management strategy, but they put research co-ordinators in the difficult position of needing to distinguish mission-creep, which must be discouraged with a negative feedback mechanism, from potential innovations, which we need to reinforce.

Every working model has a stable theoretical domain that usually represents a whole genus of similar systems and a hypothetical domain that represents a particular substantive system of interest. The theoretical component exists in some symbolic space and has usually been validated in the conventional sense of the word - it has been modified and aligned to fit past experience. The hypothetical domain, of course, is substantive, context-specific and conjectural and, as such cannot be validated historically. The theoretical domain of a model can be the intellectual property of small group of co-operating specialists as is the case with the monograph on the Behavioural Ecology of Projects prepared for COMPLEX, or it can form the foundation on which a whole discipline or institutional structure stands. In the latter case, the disciplines and institutions whose survival depends on those theories are likely to reject any argument that undermines it. The theory that science-based RTD drives innovation and economic development, for example, is culturally embedded in DG Research, and projects are unlikely to be selected for funding that explicitly challenge it.

The method of boundary conditions described here is an extension of the ‘rational falsifiability’ approach developed by the philosopher Karl Popper[14] who even tried to hi-jack science by redefining it as the investigation of falsifiable hypotheses. Popper’s definition simply doesn’t work in policy-relevant science. When scientists began to realise the seriousness of the threat faced by global warming, for example, they didn’t log the prediction in an obscure journal and start gathering the data that might falsify it, they began a programme of advocacy intended to trigger innovations that would change the predicted course of history. The political furore that followed about whether the evidence was strong enough to justify action was, in part, the product of a power-struggle between institutional actors and citizens and, in part, a stubborn insistence on applying late 19th century ideas about the relationship between science and society to the later 20th century science of complex systems.

Although Popper was clearly wrong when he argued that science always deals with falsifiable hypotheses, he was right to assert that individual scientists tend to take the mismatch between hypothesis and evidence seriously. Many of the inconsistencies in Popper’s argument can be resolved by taking a special interest in the theoretical aspects of scientific work. Scientists acquire theoretical knowledge by a process of cultural osmosis and tend to carry that knowledge unexamined throughout their careers. They often become frustrated with ‘philosophers’ who challenge those theories. ‘Philosophers’, in this derogatory sense, are theoreticians who make practitioners uncomfortable.

Thomas Kuhn’s seminal work on the structure of scientific revolutions has given us a slightly confusing term, *paradigm*, to describe this process. In his early work, Kuhn seems to have believed that scientific paradigms are pairs of rival theories - the etymology of the word suggests things laid out side-by-side. Later work by interpreters of Kuhnian theory uses the word ‘paradigm’ to describe a theoretical worldview and to speak about a tension between rival ‘paradigms’. Either way, it is clear that revolutionary science often involves a process of generalisation that shifts emphasis from a ‘special’ theory that only really works in a narrow range of contexts to a more complex, ‘general’ theory that includes the older theory as a special case.

Over the years it has become clear that the situation in the life sciences and humanities is more complex than that in physics or chemistry. Darwin’s evolutionary theory, for example, was much more complex and general than the ‘modern synthesis’ of genetics and Darwinism that dominated the second half of the 20th century. Throughout that period there has been a lively debate about whether the modern synthesis was really a narrow, special theory and there was a need for a more generalised extension. That debate will probably continue until biologists run out of non-trivial problems that can be solved using the modern synthesis and a paradigmatic revolution will favour the emergence of a more generalised or extended synthesis[15].

Thomas Kuhn will surely be remembered for pointing out that the theoretical aspects of a discipline’s knowledge-base are deeply resistant to change and are often retained long after their ‘best-before’ date has expired. Scientific revolutions tend to occur when younger scientists lose patience with those theoretical weaknesses and institutions are too weak to resist the demand for reform. In practice these innovation-cascades are often correlated across political and social domains, creating a stick-slip dynamic characterised by phoenix-cycles of institutional collapse and renaissance.

These considerations require us to consider two broad types of innovation. The first occurs when boundary conditions are violated in a way that triggers small-scale, local epiphanies. These *lesser innovations* challenge domain-specific hypotheses about the substantive domain, but leave general theories unscathed. Lesser innovations can be very valuable and significant, with knock-on effects that result in substantial changes to the ways we manage the real-world systems of interest. The COMPLEX project, for example, has successfully participated in small-scale, context-specific innovations in the management of intermittency problems in renewable energy generation and in

[14] *(1959) The Logic of Scientific Discovery*

[15] *COMPLEX has published in this field, see: DOI: 10.3109/03014460.2014.922613 and dx.doi.org/10.11141/ia.40.3*

regionally focussed case-studies in Spain and the Netherlands. Lesser innovations are generally very welcome among project stakeholders and are usually applauded as evidence of high impact research.

The second type of innovation occurs when the theoretical, or symbolic domain is tested and found wanting. Theoretical models have been extensively validated and are stable conceptual structures reinforced by custom and habit. Often there are profound conflicts of interest, and influential players will try to explain the evidence away or discredit prospective innovators. Innovators who punch their way through these vetoes and successfully challenge well-validated theory may trigger cascades of innovations that effectively undermine institutional power-structures, generating irreversible changes in many socio-natural systems at the same time. The economic collapse of 2008, for example, undermined institutional theories about the trade in risk, money supply and economic management. At the time of writing, institutions with strong vested interests in those theories are mounting a strong rear-guard action designed to 'restructure' conventional financial and economic models, but political and scientific opposition is building. An economic innovation-cascade seems increasingly likely.

6.1 Phoenix-Cycle or Pre-Emptive Innovation?

This report is intended to be read on two levels; first as a description of a practical reflexive case-study that used the research team as a test-bed for experimental research on modelling and innovation-management, and second, as a description of a general class of innovation-related problems that block progress towards a sustainable, low-carbon economy and place institutions in a bind. An institution that has become locked-in to a particular theoretical model can be a powerful obstacle to innovation. Key institutional actors feel beleaguered and threatened by outsiders who see those institutions as inert, inept and dangerous. Neither will feel secure until a blocked innovation-cascade has been released, and that sense of insecurity reinforces receptivity boundaries and reduces innovative wriggle-room.

In these circumstances, almost the last thing institutional actors want to try is the crazy experiment of selectively relaxing regulatory controls and allowing people to fail creatively. All their instincts tell them to tighten the regulatory screws, maintain stable institutional control and get through the situation. They can't innovate till they feel safe and won't feel safe until they have innovated and so they are locked in to the hierarchical attractor that generates resentment and leads to a vicious cycle of failure and low morale.

In a complex problem-domain one has to take proper account of the histories that could have happened, but did not. Some of those possible histories would have been contingent on knowledge we did not develop and patterns of resource exploitation we did not explore. They would have been qualitatively unlike the history we are familiar with in ways we cannot even imagine. Indeed, one could argue that a simulation model that slavishly tracks one, and only one history - the history that actually happened - has been invalidated in the sense that it is constitutionally incapable of representing non-linear system flips, innovation and irreversibility.

The COMPLEX approach to innovation-management has shown itself to be capable of facilitating some of the lesser innovations that challenge the hypothetical domain of a systems model, but has so far had no measurable impact on general theory. Scientific disciplines and institutional stakeholders have the ability to block and veto findings that undermine their own stability and conventional wisdom. However there have been occasions in recent decades when institutional receptivity barriers crumbled and innovation-cascades occurred that changed the course of history irreversibly.

Sometimes these innovation-cascades arose through a peaceful shift in consensus, as one generation of powerful individuals died and was replaced by another. On other occasions geo-political catastrophes, wars and economic crises brought powerful institutions to the point of collapse. Citizens became frustrated with the status quo and a cascade of stalled innovations were forced through. Some, but by no means all of these innovations were scientifically significant. Cause-effect relations in these phoenix cycles are clearly complex. Institutions that, in periods of geo-political and economic stability were able to constrain individual agency became weakened and were eventually forced to give ground by a rising tide of small-scale demands for reform. Dissidents and criminals sometimes became establishment figures as institutional checks and balances were swept away. Often there was an element of vengeful ‘score-settling’ and many instances of injustice and unforeseen consequences.

Great phoenix-cycles seem to run on a ragged, 50-year, Kondratiev period. The dreadful recession that followed the end of the Napoleonic wars and culminated in the failed ‘birth of nations’ revolutions of the late 1840s led, in the 1860s, to the emergence of the modern nation-states, democratic reforms and an innovation-cascade that gave us Darwinian theory and modern social science. The ‘Roaring 20s’ at the end of the Great War gave us near-equilibrium ecology and universal suffrage and allowed quantum uncertainty to make it onto the university curriculum. The ‘Swinging 60s’, after WWII, was the period when the study of ecological complexity became scientifically respectable. If that Kondratieff pattern is maintained, then Europe should be right in the middle of a major geo-political catastrophe that will shortly lead to a scientific renaissance.

There are certainly parallels between the situation in Europe today and that in the run-up to the two world wars of the 20th century. Although institutional economists claim we are back in growth, the poorest and middle classes are not experiencing the benefits of this growth. Supra-national institutions and trade agreements are increasingly viewed with suspicion as attacks on national sovereignty, and the refugee crisis Europe faces is clearly out of control. A vocal minority on internet and social media is claiming that World War III has already begun and is looking for individuals and institutions to blame for that war. Mr Tony Blair and Mr George W Bush are popular choices, as are financial institutions, political economists, politicians and civil servants.

These are the sorts of situations where the ‘rare but potentially catastrophic events’ mentioned in the opening paragraphs of this report can cause rational policies to have ‘undesirable and unforeseen consequences’. If this occurs, institutional vetoes will slip and an innovation-cascade, comparable to those of the 1860s, 1920s and 1960s is likely. Future historians will doubtless be interested to know whether 21st century institutions innovated pre-emptively to avoid the conflagration, as they did in the mid 19th century; or whether they ploughed through a complete phoenix-cycle of conflagration and renaissance, as they did twice in the 20th century. Either way the conditions for an innovation-cascade seem to have been met.