Simple questions about

Open Economics

What is it? Why do we want it?

Open Data : Why don't we have it?

When could contractual construction of a Data Commons be part of the solution?

1st Open Economics International Workshop, 17-18 December 2012

Presentation notes by Paul David

**Is OpEcon a part of Open Knowledge?**
Then, what is Knowledge? And what is its
relationship to *information* and *data*?

- **K**nowledge is the capability formed from **I**nformation

- **I**nformation is the signal(s) extracted from **D**ata -- using **K**nowledge

- **I**nformation is translated into *actions* (based on **K**nowledge), including the generation and capture of **D**ata

Knowledge as a human capability --
and the "*K I D* - triangles"

K

I   D
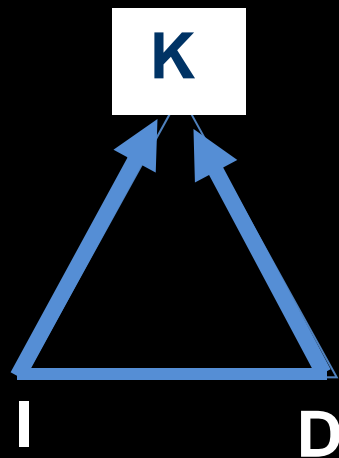
building **knowledge** *(capabilities for action)*

I

D   K
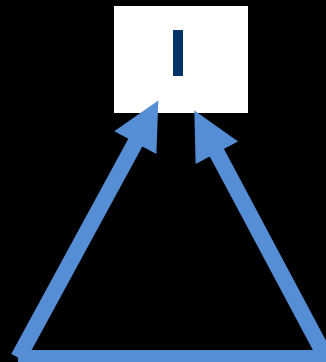
extracting **information** *("signals") from data*

D

K   I
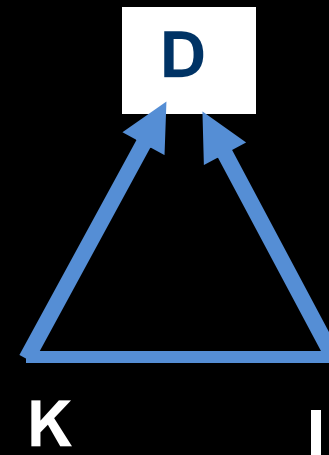
Generating and capturing *data*

Information and *disciplinary knowledge"*
Is social community knowledge the *intersection* or the *union* of the individual members' knowledge?

isolated knowledge pools

C & C

societal knowledge

Codification and Communication forms societal (common) knowledge, augmenting individual capabilities with shared "tools"
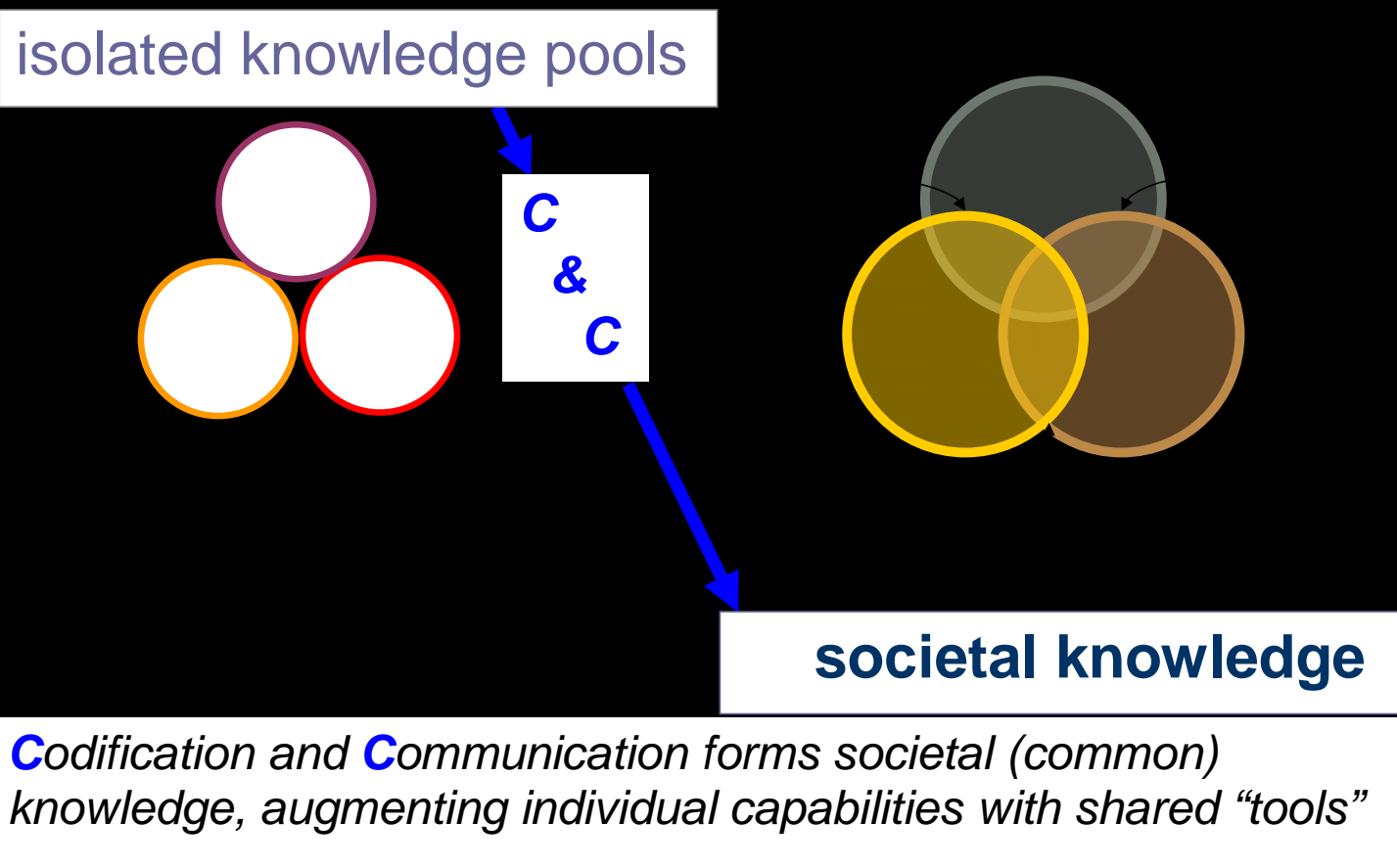
To what does "Open" refer?

o *Things that have been revealed, rather than kept secret, and hence may be included in Common Knowledge?*

o *Social practices and procedures governing the conduct of 'economic science '?*

o *Does it then refer to the "open-ness" of Open Science?*

Open science's ethos and institutions

Institutional features structure resource allocation in 'the Republic of (Open) Science'

The key institutionalized social 'norms' that R. K. Merton (1938, 1973) identified and M. Polanyi (1962) described as social system, are readily remembered using J. Ziman's (1994) mnemonic :

C ooperation

U niversalism

D isinterestedness

O pen-ness

S cepticism

Idealized social norms and institutionalized procedures of the Republic of (Open) Science

Idealized social norms

- cooperation and trust among scientists
- autonomy in determination of research agendas
- personal disinterestedness in research outcomes
- full disclosure of findings and methods
- expectation of verification by replication

Idealized social norms and institutionalized procedures of the Republic of (Open) Science

Stylized procedural arrangements

- rewards based upon collegiate reputational status

- reputation based on peer-appraisal of 'scientific contributions'

- eligibility for evaluation based upon non-ascriptive characteristics

- substantial autonom of individual in design and research conduct is expected (and with this goes responsibliity for the research)

- 'a scientific contribution' requires validation of the researcher's claim to priority in discovery or invention

Institutional features and resource allocation in the Republic of (Open) Science

Functionality of the cooperative system that promotes rapid cumulative growth in reliable knowledge:

- "collegiate reputational reward structure" (CRRS) provides incentives and signals for agents' effort allocation decisions

- scientific reputational standing is based on community acknowledgement of claims to "priority of discovery"

- incentive compatibility of priority with the norm of *open-ness* (full disclosure) :

  asymmetric information → problems of input monitoring → monitoring of output, with rewards for priority → rapid disclosure

- *disclosure, skepticism and disinterestedness* → validation of results →promotes rapid "closure" (effective consensus)

- *universalism* prevents "homogenization" of social communications network structure → protecting deviant opinion from premature formation of consensus (dogmatic belief)

The "Logical Origins" of Open-Science Institutions:
A Functionalist Rationale

GOAL: Rapid Accumulation of Reliable Knowledge

Public Knowledge

Validation

Disclosure Norms

Publication Procedures

Priority Races

Reward System
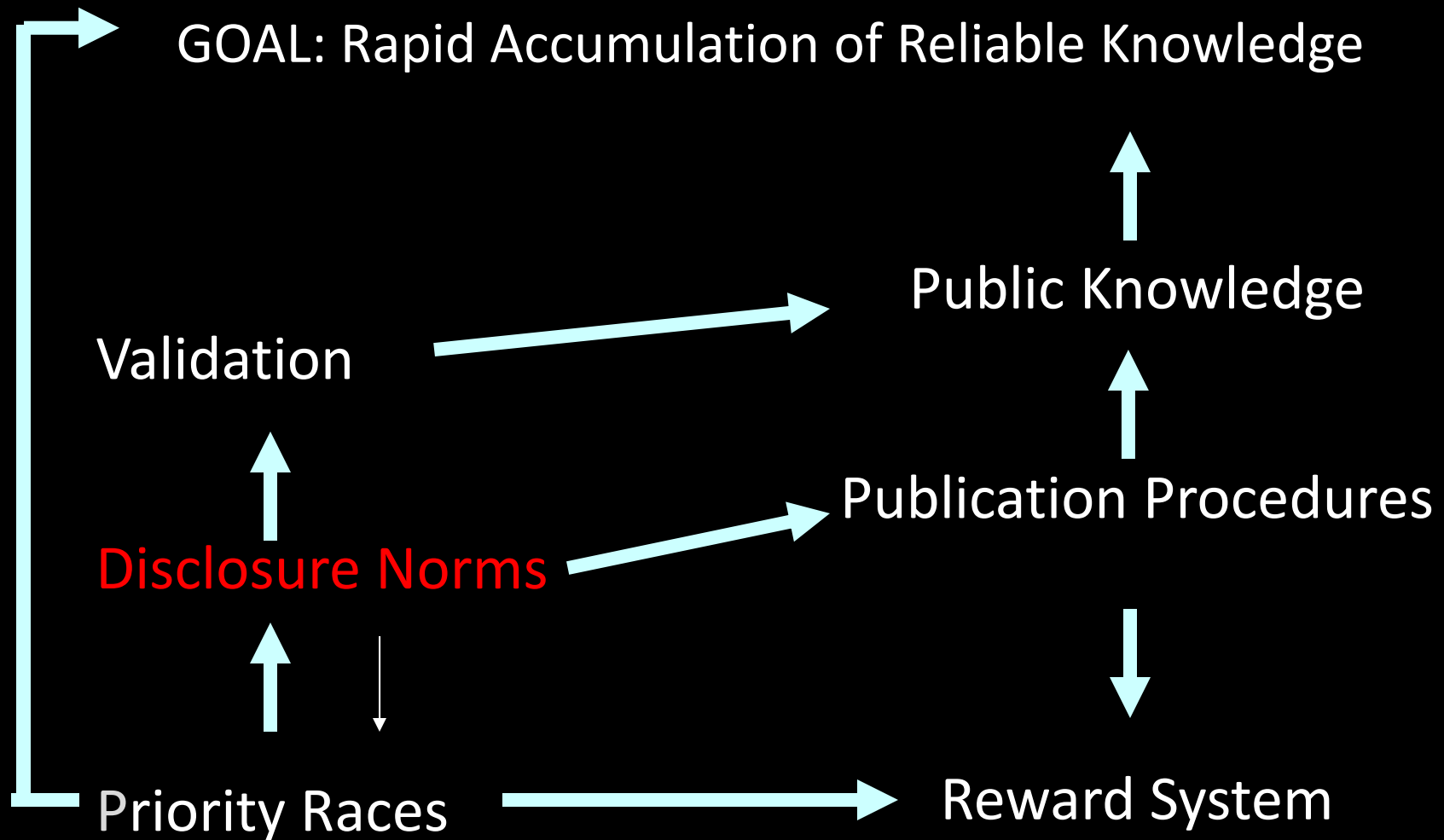
Open data – as a resource for others

- clearly documented provenance

- conceptually appropriate mode of initial creation/observation

- identification of filtering and editing rules

- unambiguous terms of use : copyleft?

Providing others with "good datasets" is costly for those who best know the data and its likely potential uses – those involved in the research projects that created the data. But as a rule, they can ill afford to undertake the expense.

SOME DIFFICULTIES ENSUE

- Publicly funded projects rarely provide support for the data editing and documentation activities –esp. in social sciences.

- Systematic recording of steps in the data-generation process, and subsequent documentation are time-consuming and low priority tasks that typically are deferred to end of project – when funds have been expended

- The open science reward system has no institutionalized mechanisms that incentivize careful data preservation and documentation. But, since releasing inadequately "cleaned" and poorly documented datasets is likely to create a "time sink" for investigators when external users find errors, ambiguities and ask for help to use the dataset: doing nothing seems best.

- Incentive mechanisms would need to begin by defining "data quality dimensions, setting minimum standards, and specifying performance metrics –which will be heterogeneous, as they must be appropriate for the research area of intended use.

THE ANATOMY OF THE "RESEARCH DATA ANTI-COMMONS"

Layer 1: Search costs, …to discover whether tools described in the research literature are privately appropriated, and to whom the property rights were assigned, whether as patents, or as copyright computer code, or as database rights.

Layer 2: Transactions costs, …. strictly these arise when one has identified the owner(s) of the IPR and seeks a license, or an agreement to transfer materials

THE ANATOMY OF THE "RESEARCH DATA ANTI-COMMONS", contd

Layer 3: Multiple-marginalization and royalty-stacking... Even when there are no strategic "hold-outs", the distribution of exclusion rights to multiple items means that access to each may be "priced" (or otherwise impose a cost on the recipient) without regard for the effect of the resulting negative pecuniary externalities that impact the demand for other items.

When tools are *gross complements*, rather than substitutes, the resulting inefficiency is the dual of the that produced by ignoring congestions externalities. Here the unit access costs ("prices") of components ignores the pecuniary externalities on the demand for the project as a whole, resulting sub-optimal use of the entire bundle.

The severity of the inefficiency increases with the number of tools that are strict complements for the given research project.

"RESEARCH DATA ANTI-COMMONS"– Where is the incidence of damage?

Multiple-marginalization should be seen not only as potentially impeding the use of patented or copyrighted research tools, and thereby blocking formation of some research projects, but more generally as degrading the exploration of large data-fields – or "discovery spaces" – which are now particularly important in exploratory research, and likely to become moreso in the epoch of Big Data.

Consider a simple model of a research production project: the output is results R, produced under cost-minimizing conditions on a budget of G

$$G = \sum [p_{\{i\}}] [b_{\{i\}}] + X,$$

according to production function

$$R = F(S , X),$$

where

X is a vector of inputs of experimental time and equipment

and

S is the output of a *search* activity, according to search function:

$$S = S (b_{\{1\}}, b_{\{2\}}, ...b_{\{B\}}), \text{ in which}$$

$b_{\{i\}}$ is the information extracted from database i.

"RESEARCH DATA ANTI-COMMONS"– Modelng the incidence of damage

Notes for economist who make care about the modelling steps:

1) For simplicity, symmetry of intensity of database use is assumed, and all projects are also assumed to have identical search strategies.

2) From a CES production function for "search" one obtains derived demands for access to database contents, as a function of unit extraction "charges", project real budget level and the elasticity of substitution among databases.

3) Assume database holders independently set profit maximizing royalty rates (or, equivalently, impose "own surplus -maximizing " reductions in the quality of data) that others extract (receive) from their database -- as discriminating monopolists would do). Assume for simplicity that database holders are symmetric, and solve for the resulting relative (equivalent) charges that will be set, and for and the research project's consequent cost-minimizing searchs in the data thus obtained, and its production decisions.

Basic solution results: Even if the b {i} are not strict complements, and there is symmetric non-zero elasticity of substitution between them, when database rights are separately owned and "priced" individually (without taking account of pecuniary spillovers) so as to maximize the owners' separate revenues (or minimize their costs with a fixed budget), then: the larger the number of databases, B, the more severely degraded will be S. Hence R (research output) for given project funding levels will be reduced – so long as S and X are not infinitely substitutable. *The outcome is inferior to that obtained with joint monopoly ownership of databases.*

The contractually constructed quasi-commons (or "club commons") is the immediately feasible remedy for the anti-commons --
and also for other less serious barriers to collaborative production of information and data resources.
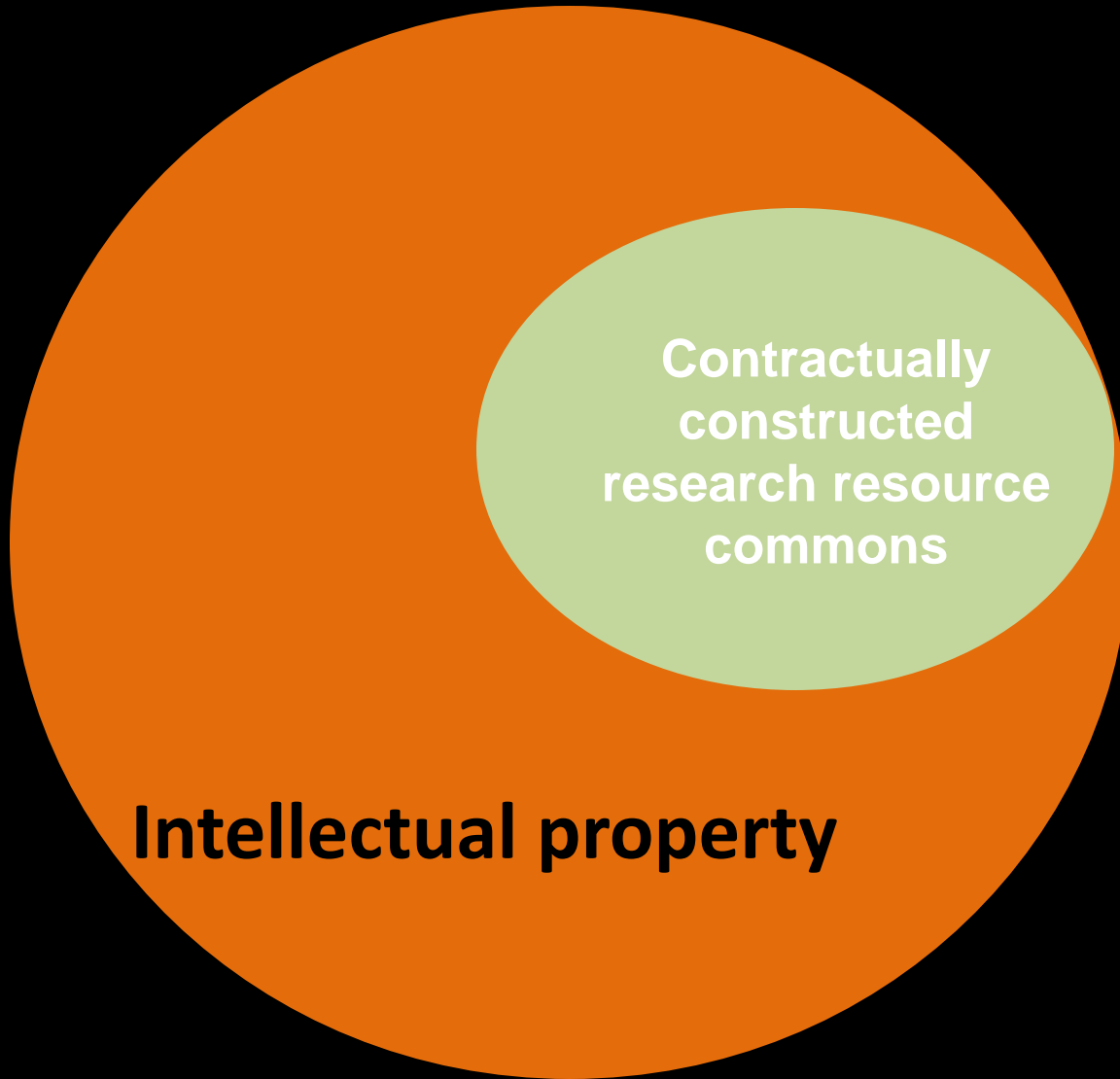
   – It can make use of the legal protection afforded by the IPR regime, and the latter's limitations on total and indefinite monopoly ownership; IPR is this context is better than monopoly enforced by "self-help" (encryption technology).
   -- It utilizes contract law to enforce compliance with voluntarity entered agreements to pool content under common use or other cross-licensing and "sharing" arrangments among members of the commons.

An example from the extensive historical experience of well-*governed* and *managed* Common Property Resources: there was NO "Tragedy of the Medieval Village Commons"!

The terrier of Salford Manor, in Oxfordshire records the following two items among the by-laws adopted by common consent of the "inhabitants" on 17th September, 1592:

"1. *Imprimis* it is agreed that every inhabitant may kepe for every three acres of follow [fallow]that he hath within this parryssh eight sheepe and not above upon payne for every sheepe he shall kepe above that rate to foryte every tyme xij d [12 pence, i.e. one shilling]";

"7. Item that every may kepe for every five acres of land in one field [referring to the three open-fields of the arable land in the village] foure kyne [kine referring to 'cows'] and not above upon payne of iij s. iij d. [3 shillings and 3 pence] ."

Source: Salford Manor,No.368, in the Codrington Library (All Souls College, Oxford), transcribed and printed as doc. 216 in Ault (1965: Appendix, p. 93).