towards a "data web"

"raw data now"

source: w3.org

rdf + owl
open source software
open licensing

The Systemic Value of Compatibly
Communicating Devices Grows as the
Square of Their Number:

reality check.

Technical Specifications

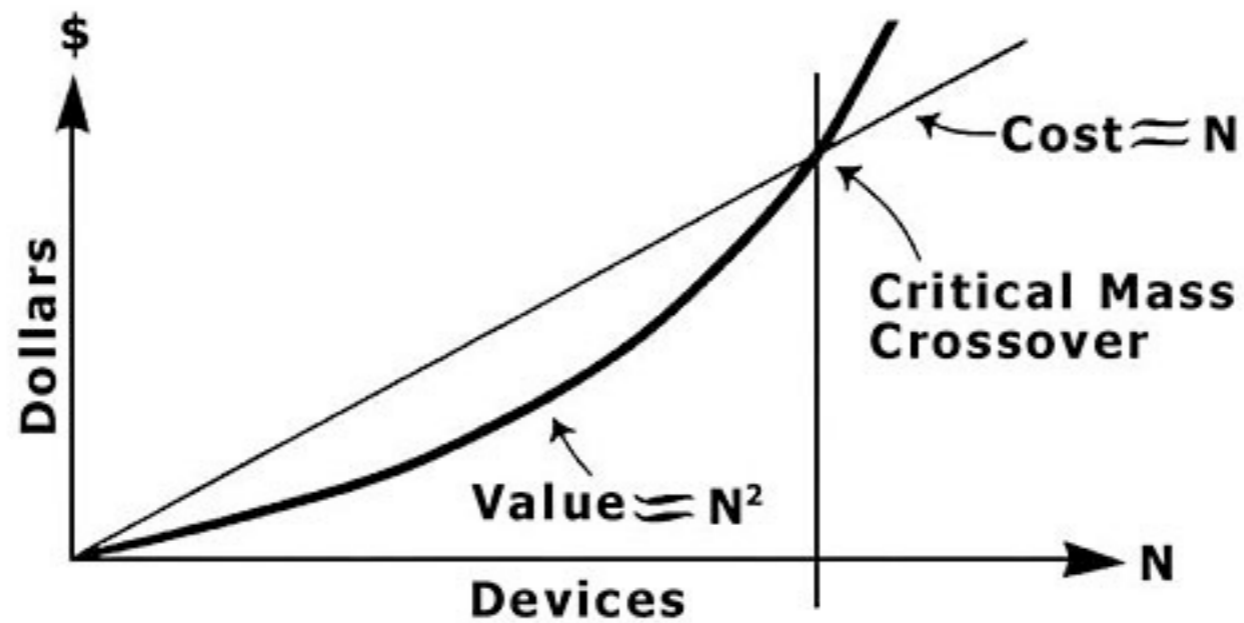| Group	Title	Most stable	In progress	Version history
App	Simple Application Messaging Protocol	1.11		1.11 1.11 1.10 1.00
DAL	Simple Cone Search	1.03		1.03 1.02 1.01 1.00
	Simple Image Access	1.0		1.0 1.0 1.0 1.01 1.00
	Simple Line Access	1.0	RFC	1.0 1.0 1.0
	Simple Spectral Access	1.04		1.04 1.03 1.02 1.01 1.01 1.00
	Table Access Protocol	1.0		1.0 1.0 1.0 1.0 1.00
	IVOA Astronomical Data Query Language	2.00		2.00 2.00 2.00 1.01 1.00
	IVOA SkyNode Interface	1.01		1.01 1.00
DaM	Space-Time Coordinate Metadata for the Virtual Observatory (STC)	1.33		1.33 1.31 1.30 1.21 1.20 1.10 1.00
	Data Model for Astronomical DataSet Characterisation	1.13		1.13 1.12 1.12 1.11 1.10 1.00
	Simple Spectral Lines Data Model	1.0	RFC	1.0 1.0 1.0
	IVOA Spectral Data Model	1.03		1.03 1.02 1.01 1.01 1.00
GWS	IVOA Single-Sign-On Profile: Authentication Mechanisms	1.01		1.01 1.01 1.00 1.00
	VOSpace service specification	1.15	2.0	2.0 2.0 1.15 2.0 1.15 1.15 1.14 1.13 1.12 1.12 1.11 1.10 1.02 1.02 1.01 1.00 1.00
	IVOA Credential Delegation Protocol	1.0		1.0 1.0 1.01 1.01 1.00
	Universal Worker Service	1.0	RFC	1.0 1.0 1.0
	IVOA Support Interfaces	1.0	RFC	1.0 1.0 1.0
	IVOA Web Service Basic Profile	1.0	RFC	1.0 1.0
ReR	IVOA Identifiers	1.12		1.12 1.11 1.10 1.10 1.10 1.00
	IVOA Registry Interfaces	1.0		1.0 1.0 1.00 1.02 1.01 1.00
	Resource Metadata for the Virtual Observatory	1.12		1.12 1.12 1.10 1.10 1.01 1.01 1.00 1.00
	StandardsRegExt: a VOResource Schema Extension for Describing IVOA Standards	1.0		1.0
	VOResource: an XML Encoding Schema for Resource Metadata	1.03		1.03 1.02 1.02 1.01 1.00
	VODataService: a VOResource Schema Extension for Describing Collections and Services	1.1	RFC	1.1 1.1 1.10
Semantics	An IVOA standard for Unified Content Descriptors	1.10		1.10 1.10 1.06 1.05 1.03
	UCD1+ Controlled Vocabulary	1.23		1.23 1.22 1.21 1.20 1.20 1.11 1.11 1.10 1.02 1.00
	Maintenance of the list of UCD words	1.20		1.20 1.20 1.10 1.00
	Vocabularies in the Virtual Observatory	1.19		1.19 1.18 1.16 1.15 1.13 1.00
SDP	IVOA Document Standards	1.2		1.2 1.2 1.2 1.2 1.2 1.1 1.1 1.0 1.0
VOE	Sky Event Reporting Metadata (VOEvent)	1.11		1.11 1.11 1.10 1.01
VOT	VOTable Format Specification	1.2		1.2 1.2 1.2 1.20 1.20 1.10 1.00

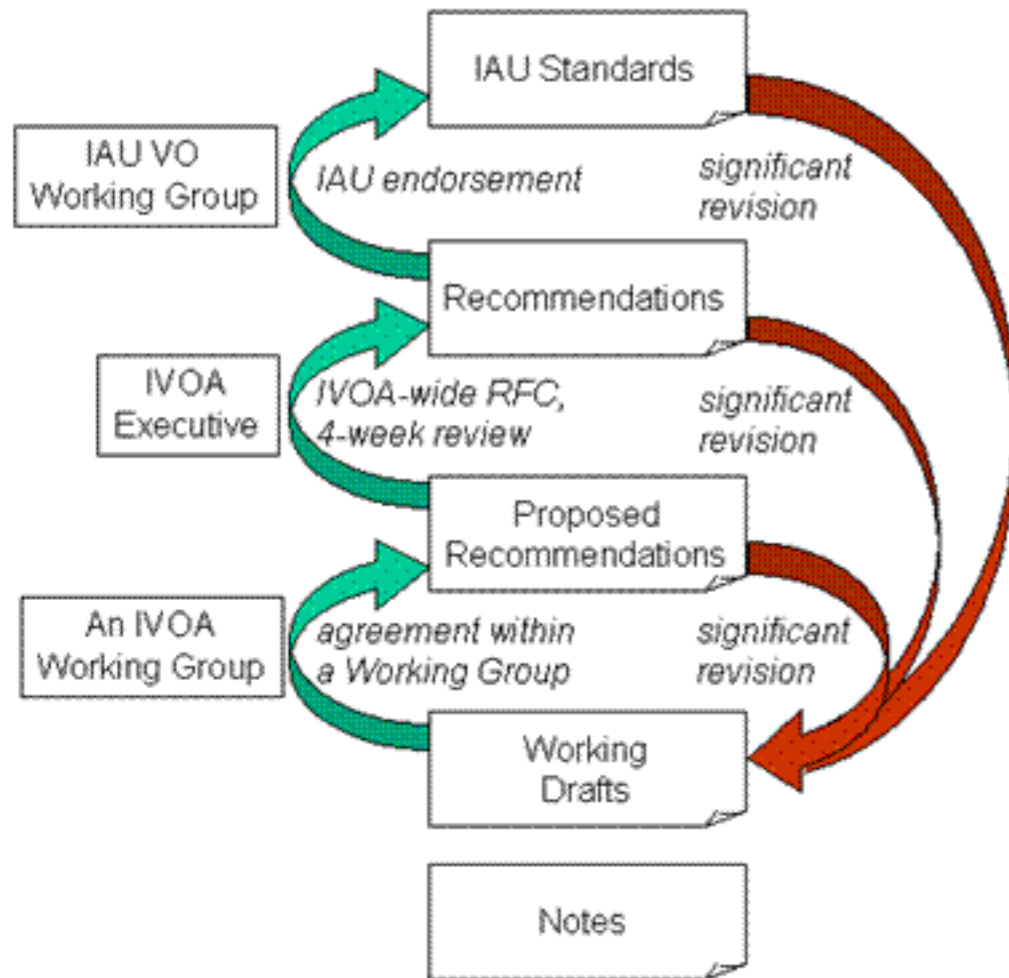Maturity level: Recommendation Proposed Recommendation Working Draft

Notes

Title	Latest Version
IVOA Procedures	
Guidelines for Participation	1.00
Software Licensing Guidelines	1.00
Document Standards Management: Guidelines and Procedures	1.00
General	
Charter for the IVOA Technical Coordination Group	1.1
Design Notes for Revised IVOA Web Pages	1.00
The IVOA in 2009: Technical Assessment and Roadmap	1.00
The IVOA in 2008: Technical Assessment and Roadmap	1.00
The IVOA in 2007: Assessment and Future Roadmap	1.00
The IVOA in 2006: Assessment and Future Roadmap	1.00
The IVOA in 2005: Assessment and Future Roadmap	1.00
Report to the 24th CODATA General Assembly on data activities in the IAU	1.00
Management, Storage, and Utilization of Astronomical Data in the 21st Century	1.00
Virtual Observatory Architecture Overview	1.00
Technical	
Referencing STC in VOTable	2.0
STC-S	1.33
DAL2 Service Architecture and Standard Profile	1.0
Extending VOEvent for more complex data	1.0
Ontology of Astronomical Object Types Use Cases	1.1
Ontology of Astronomical Object Types	1.3
Dakota: A Cross-Platform VOEvent Broker	1.0
Note on the description of polarization data	1.0
VOEvent Transport Protocol	1.1
An encoding system to represent stellar spectral classes in archival databases and catalogs	1.03
Reflections on VOEvent 1.1 and the Real World	1.00
Enabling a robust VOSpace for VO tools and services	1.10
S3: Proposal for a simple protocol to handle theoretical data (microsimulations)	1.00
Proposal for a Simulation Database Standard	1.00
A Proposal for Digital Signatures in VOEvent Messages	1.10
Space-Time Coordinate (STC) Metadata Model	1.31
Introduction to CEA and UWS	1.00
An RDF version of the VO Registry	1.00
Implementations of the Characterisation Data Model	1.12
Footprint Overlay Specification	1.01
The Registry of Registries	1.00
Utype list for the Characterisation Data Model	1.11
Astronomical Keywords in the era of the Virtual Observatory	1.00
UTypes and URIs	1.00
UWS recast as a REST protocol	1.00
PLASTIC - a protocol for desktop application interoperability	1.00
Ontology of astronomical object types [1]	1.00
Astronomical Outreach Imagery Metadata Tags for the Virtual Observatory	1.00
Ranking Query Result Sets	1.00

ontology of astronomical object types

IVOA Document Standards Process

IAU VO Working Group — IAU Standards

IAU endorsement — significant revision

IVOA Executive — Recommendations

IVOA-wide RFC, 4-week review — significant revision

An IVOA Working Group — Proposed Recommendations

agreement within a Working Group — significant revision

Working Drafts

Notes

Submission Log

Date	Document ID	Version	Initiator	I/W Group	Action
2010-06-23	VOTableSTC-20100618	2.0	Markus Demleitner	VOTableSTC	updated
2010-06-17	StandardsRegExt-20100519	1.0	Ray Plante	ReR	published
2010-06-15	STC-S-20091030	1.33	A. Rots	DaM	updated
2010-05-21	DAL2Arch-20100521	1.0	Doug Tody	DAL	published
2010-05-18	DocStd-20100413	1.2	Francoise Genova	SDP	PR to REC
2010-05-11	SSLDM-20100506	1.0	Pedro Osuna	DaM	updated
2010-05-11	SLAP-20091016	1.0	Jesus Salgado	DAL	updated
2010-04-21	TAP-20100327	1.0	Patrick Dowler	DAL	PR to REC
2010-04-15	VODataService-20100412	1.1	Ray Plante	ReR	updated
2010-04-15	TCG-Charter-20100329	1.1	Christophe Arviset	N/A	updated
2010-03-24	VOSpace-20100323	2.0	Matthew Graham	GWS	updated
2010-03-24	VOEventTables-20100323	1.0	Roy Williams	VOE	published
2010-03-19	IVOAWebDesign-20100301	1.0	Andy Lawrence	N/A	published
2010-03-19	VOSI-20100311	1.0	Matthew Graham	GWS	WD to PR
2010-03-11	AstrObjectOntologyUseCases-20100117	1.1	Alexandre Richard	Semantics	updated
2010-03-04	AstrObjectOntology-20100117	1.3	Alexandre Richard	Semantics	updated
2010-03-01	WSBasicProfile-20100226	1.0	Matthew Graham	GWS	WD to PR
2010-03-01	CredentialDelegation-20100218	1.0	Matthew Graham	GWS	PR to REC
2010-02-25	TAP-20100225	1.0	Patrick Dowler	DAL	updated
2010-02-11	UWS-20100210	1.0	Paul Harrison	GWS	updated
2010-02-10	DakotaBroker-20100210	1.00	Robert B. Denny	VOE	published
2010-02-05	SIA-20091116	1.0	Paul Harrison	DAL	PR to REC
2010-02-04	Polarization-20100203	1.0	Anita Richards	DM	published
2010-01-08	TAP-20091225	1.0	Patrick Dowler	DAL	updated
2009-12-02	VOTable-20091130	1.2	Francois Ochsenbein	VOT	PR to REC
2009-11-05	RegistryInterface-20091104	1.0	Ray Plante	ReR	PR to REC
2009-11-02	VOTable-20091102	1.2	Francois Ochsenbein	VOT	updated
2009-10-15	DocStd-20091008	1.2	Robert Hanisch	SDP	updated
2009-10-12	Vocabularies-20091007	1.19	Norman Gray	Semantics	PR to REC
2009-10-12	VOSpace-20091007	1.15	Matthew Graham	GWS	PR to REC
2009-10-12	SIA-20091008	1.0	Paul Harrison	DAL	updated
2009-10-08	TAP-20091006	1.0	Patrick Dowler	DAL	updated
2009-10-02	DocStd-20091002	1.2	Robert Hanisch	SDP	updated
2009-10-02	TAP-20090608	1.0	Patrick Dowler	DAL	WD to PR
2009-10-01	IVOATechRoadmap2009-20091001	1.00	Christophe Arviset	N/A	published
2009-09-29	VOTable-20090929	1.2	Francois Ochsenbein	VOT	updated
2009-09-23	RegistryInterface-20090917	1.0	Ray Plante	ReR	updated
2009-09-23	VOSpace-20090921	1.15	Matthew Graham	GWS	updated
2009-09-16	VODataService-20090903	1.1	Ray Plante	ReR	WD to PR
2009-09-09	UWS-20090909	1.0	Paul Harrison	GWS	WD to PR
2009-09-07	CredentialDelegation-20090818	1.0	Matthew Graham	GWS	updated
2009-09-02	SSLDM-20090714	1.0	M. Louys	DaM	WD to PR
2009-09-02	SSLDM-20090713	1.0	Jesus Salgado	DaM	published
2009-09-02	SLAP-20090714	1.0	Jesus Salgado	DAL	WD to PR
2009-09-02	SLAP-20090713	1.0	Jesus Salgado	DAL	published
2009-09-01	VOEventTransport-20090805	1.1	Robert Deny	VOE	updated
2009-08-28	Vocabularies-20090825	1.18	Norman Gray	Semantics	updated

"raw metadata and standards processes, and consensus processes, and document submission standards, and archives"

"raw metadata and standards processes, and consensus processes, and document submission standards, and archives"
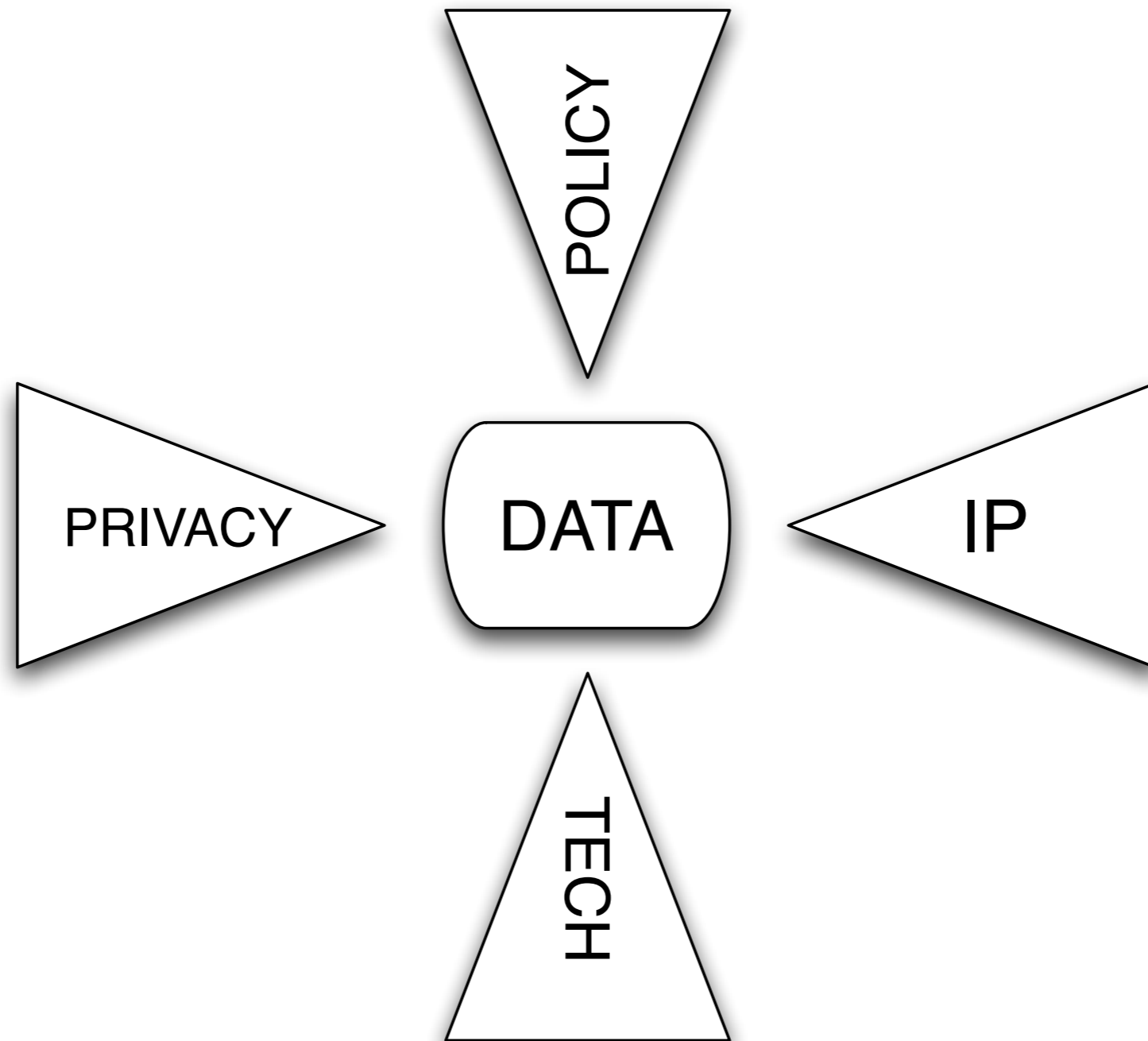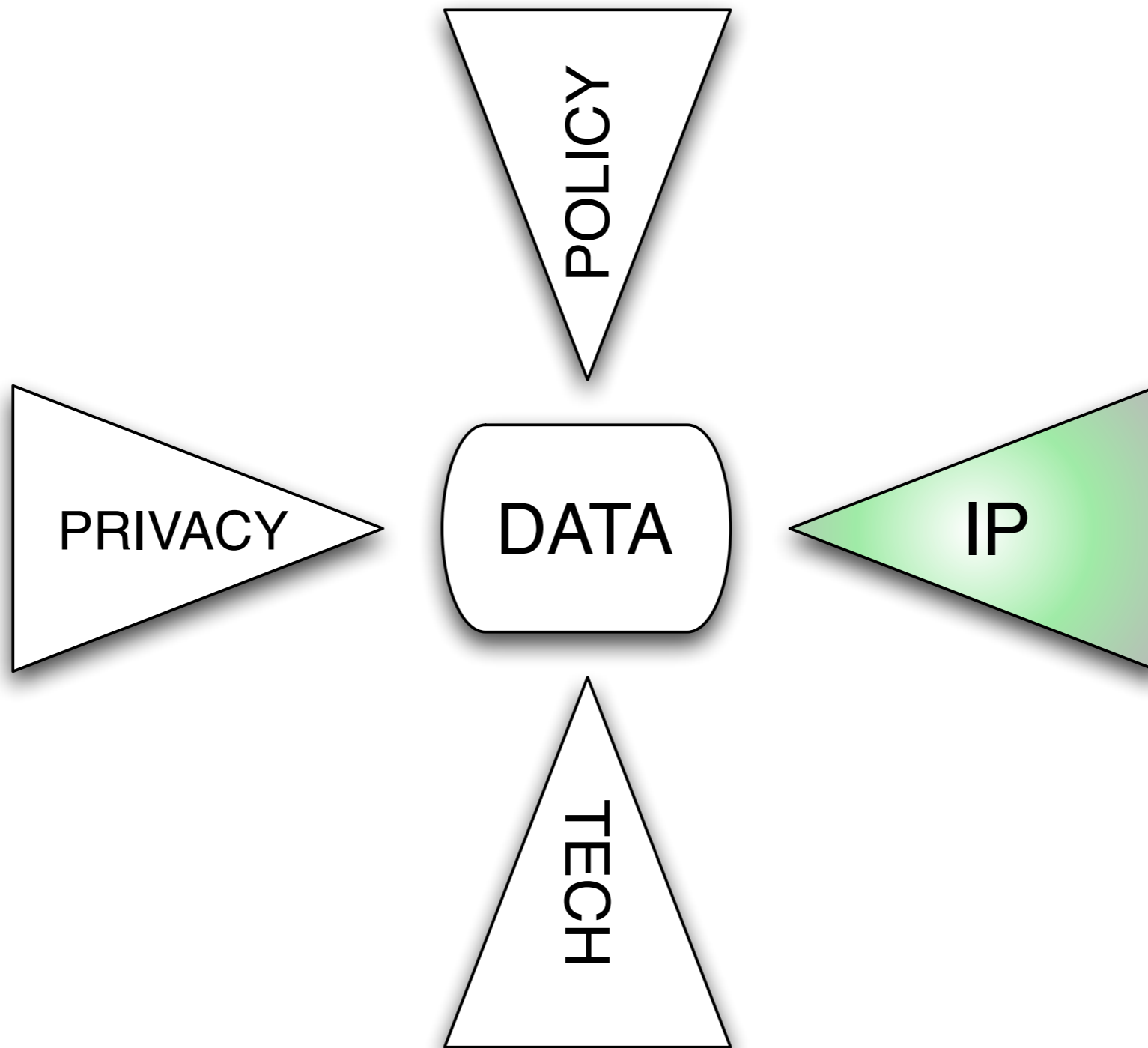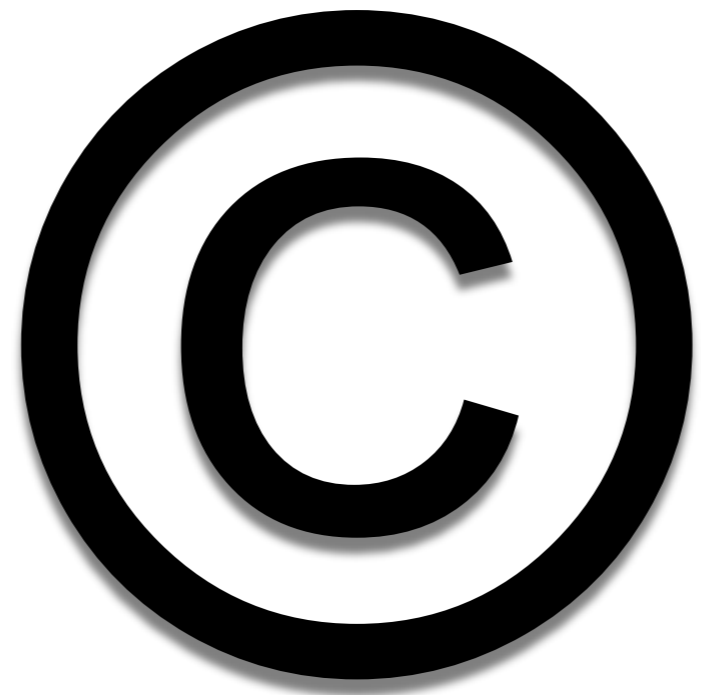
NOW!
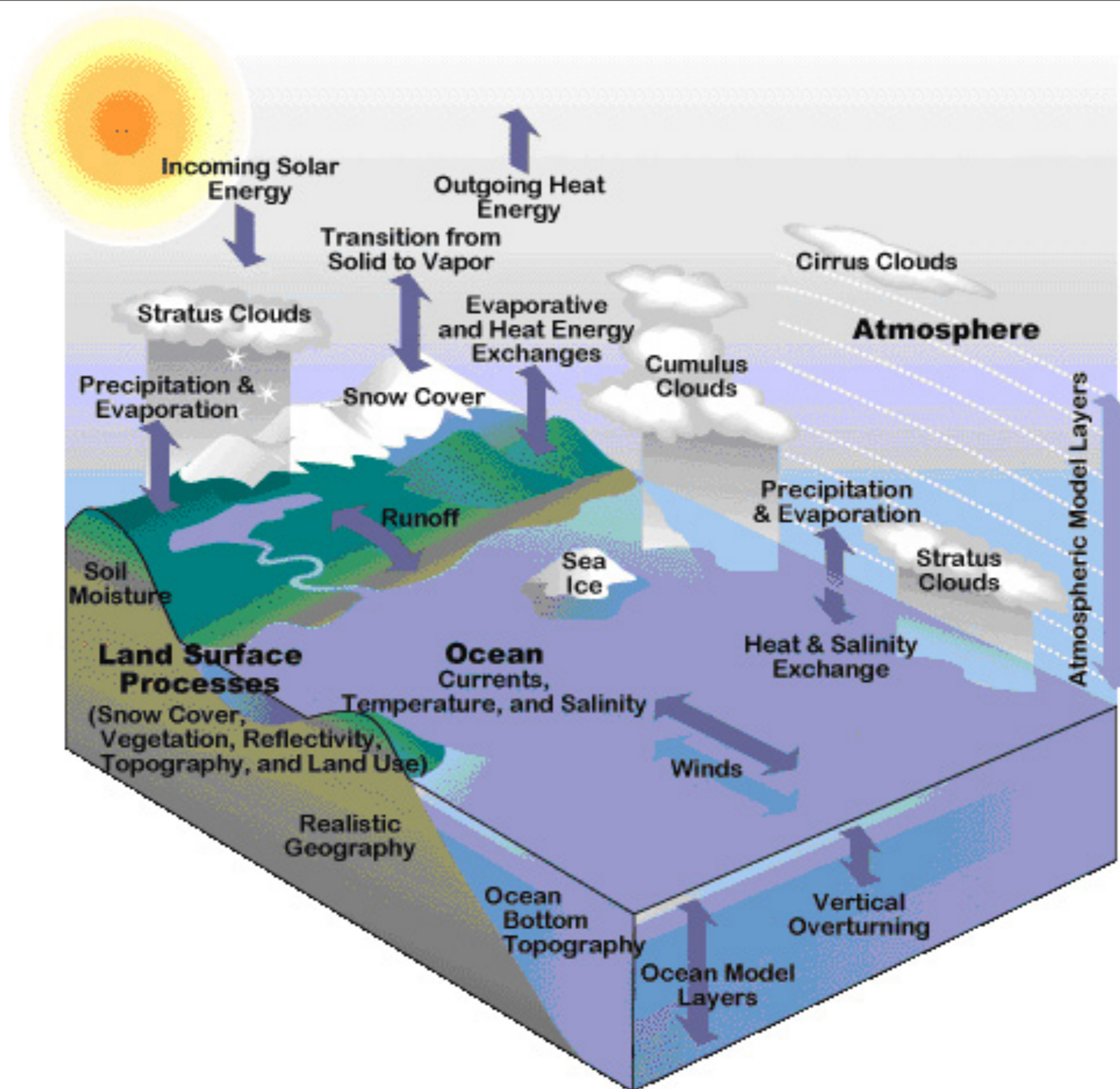
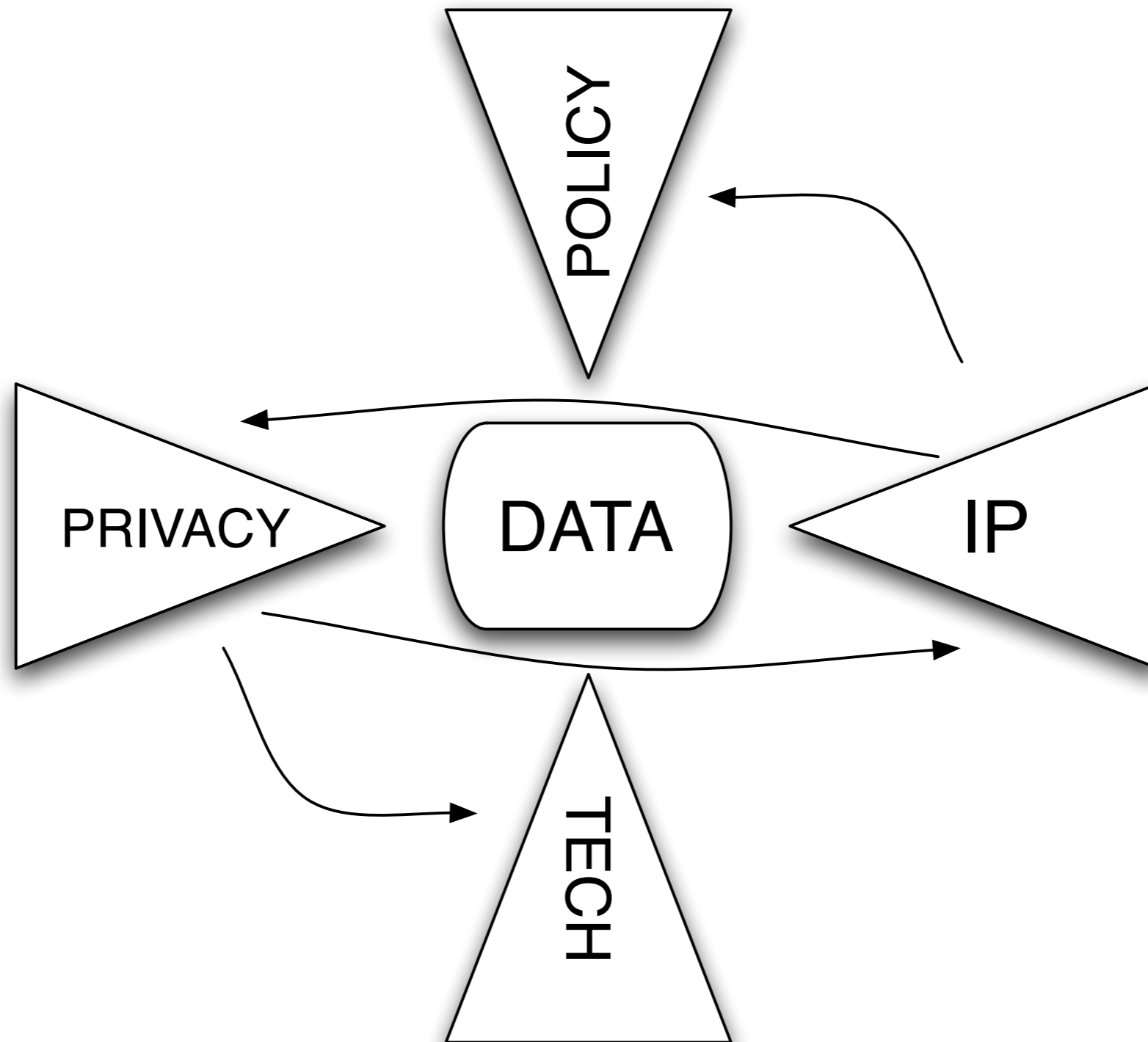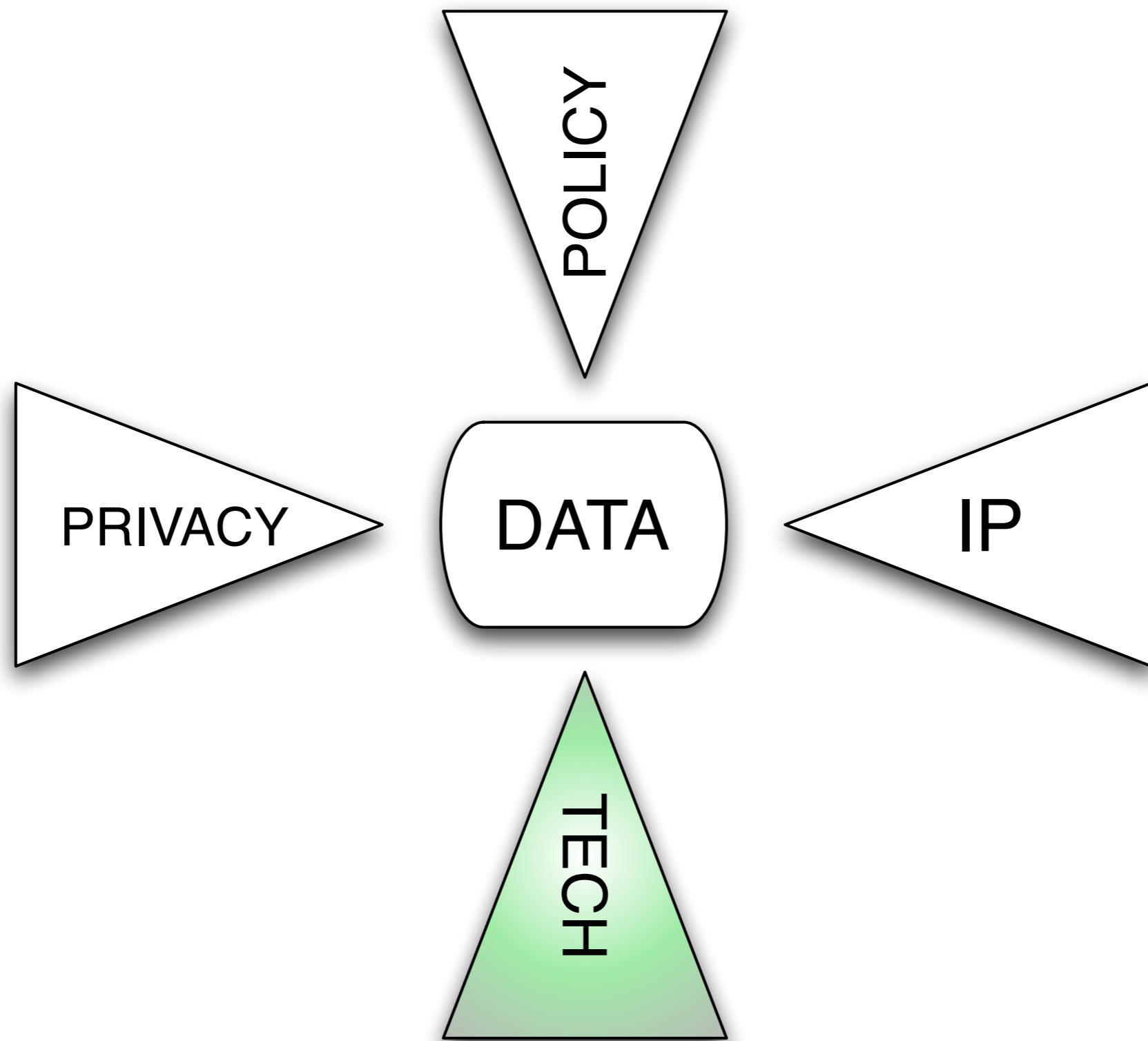gathered independently, for different reasons...

what things regulate?

# in relation to data...

IP interference...

Bundles

The Neurocommons RDF distribution is organized into modules or "bundles". Following is list of what's provided. Each has its own page of documentation.

In most cases bundle B corresponds to named graph http://purl.org/science/graph/B.

| Bundle	Description	Documentation
Derived from MeSH:		
mesh/mesh-skos	MeSH polyhierarchy represented using SKOS courtesy van Assem et al.	/mesh/mesh-skos
mesh/qualified-headings	MeSH qualified headings - defines one URI for each valid major/minor heading combination	/mesh/qualified-headings
Derived from Medline:		
medline/subject-headings	Medline: NLM MeSH subject headings for all articles	/medline/subject-headings
medline/titles-years	Medline: title and year of publication for each article	/medline/titles-years
Derived from NCBI:		
ncbi/goa	NCBI Gene Ontology annotations	/ncbi/goa
ncbi/homologene	NCBI Homologene selection	/ncbi/homologene
ncbi/gene-info	NCBI Gene gene synonyms extraction	/ncbi/gene-info
ncbi/gene-pubmed	Links from NCBI Gene to Medline	/ncbi/gene-pubmed
Ontologies:		
bams	BAMS (Brain Architecture Management System)	/bams
galen	Galen ontology	/galen
obo/all	All OBO ontologies	/obo/all
mesh-eswc06	Supporting ontology for conversion of MeSH polyhierarchy to RDF	/mesh-eswc06
nci-thesaurus	NCI thesaurus	/nci-thesaurus
sciencecommons	Ad hoc Science Commons ontology	/sciencecommons
senselab	Senselab	/senselab
skos	W3C SKOS (Simple Knowledge Organization System) ontology	/skos
pdspki	PDSP Ki	/pdspki
Other:		
addgene	Addgene plasmid catalog	/addgene
neurocommons-text	Neurocommons text processing pilot	/neurocommons-text
aba	ABA (Allen Brain Atlas)	/aba

it's going to take time.

the appropriate metaphor is the *internet*.

POLICY

PRIVACY

DATA

IP

TECH
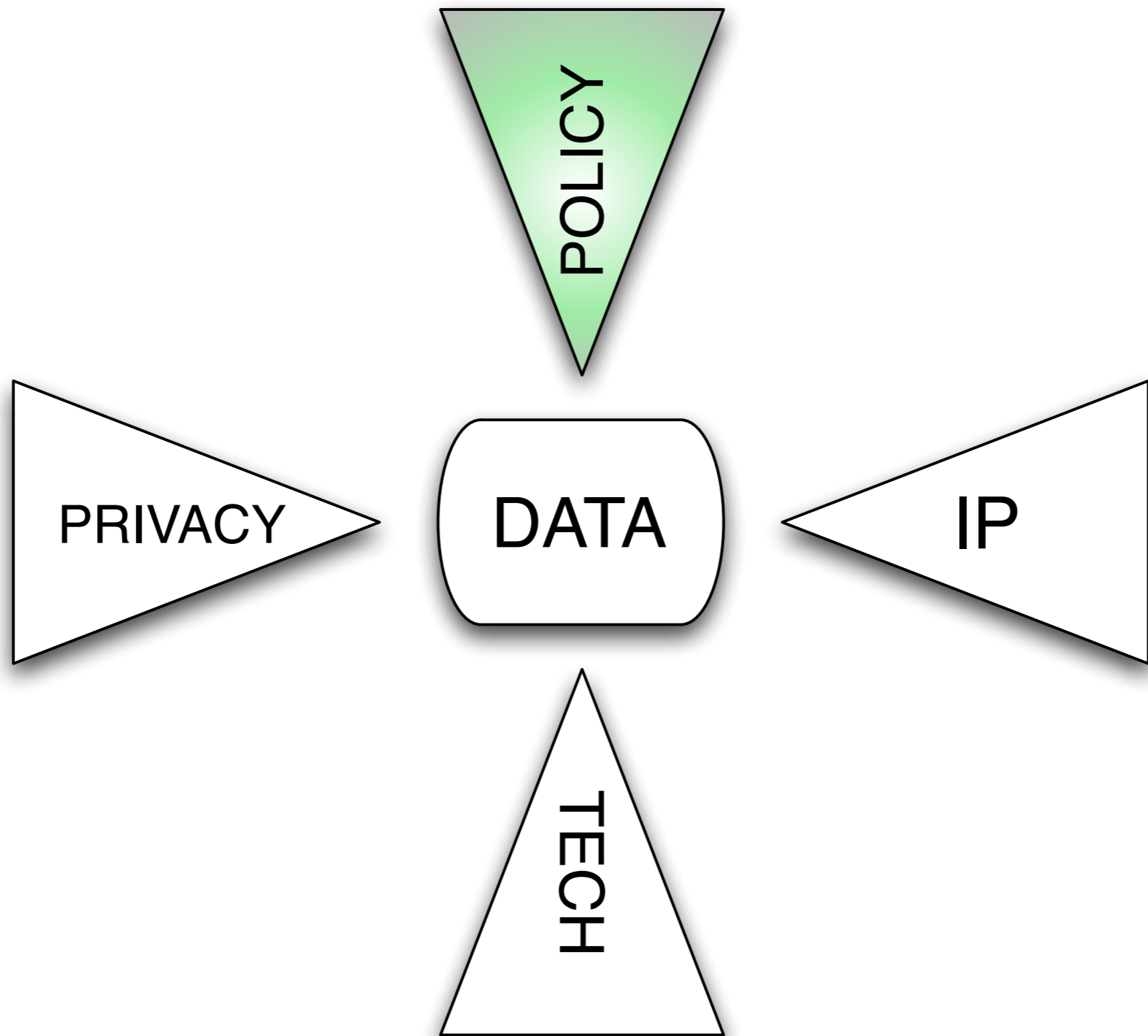
NIH reaffirms its support for the concept of data sharing. We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. The NIH endorses the sharing of final research data to serve these and other important scientific goals. The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers. Starting with the October 1, 2003 receipt date, investigators submitting an NIH application seeking $500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible.

no:
tracking
followup
accountability
transparency
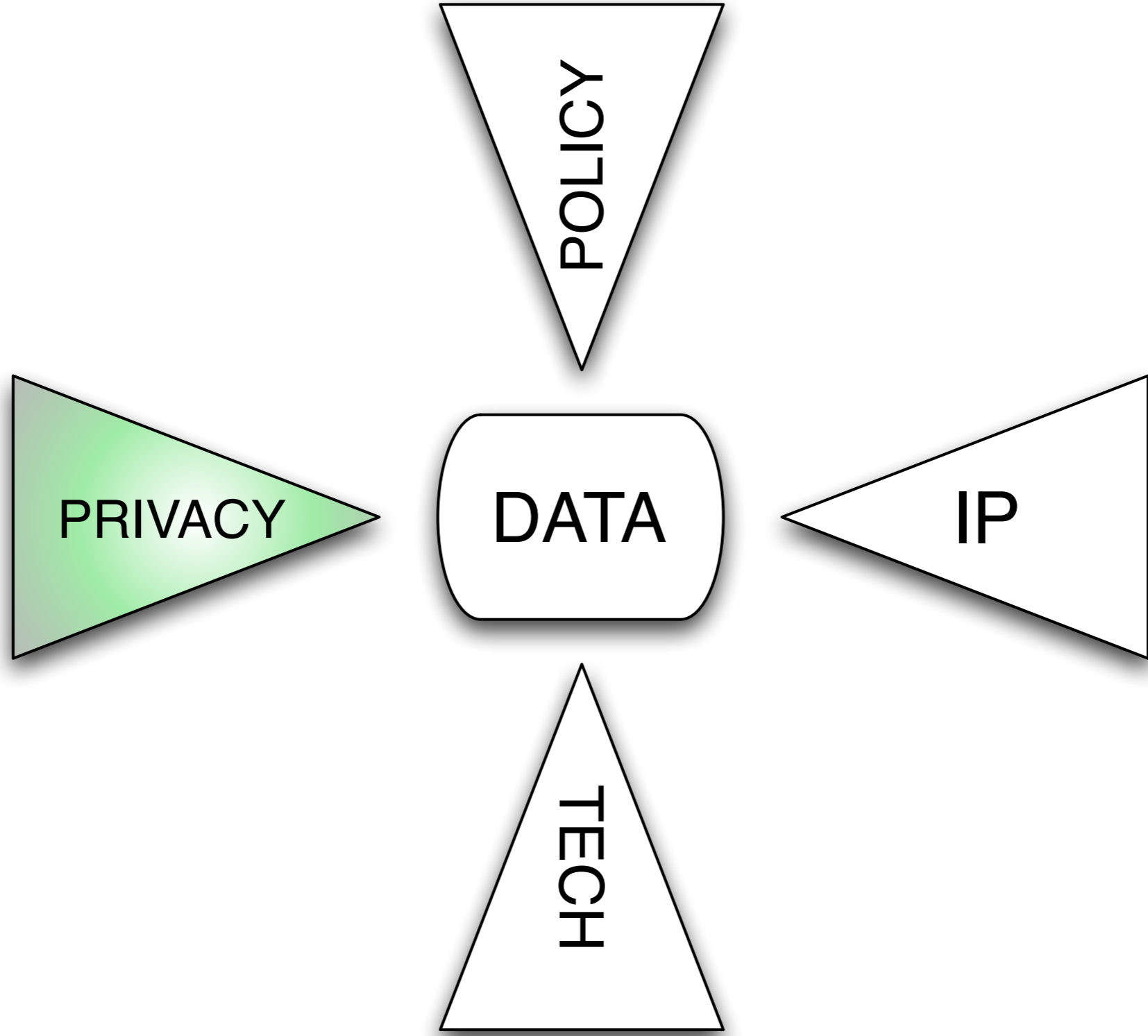
pastiche of international regimes
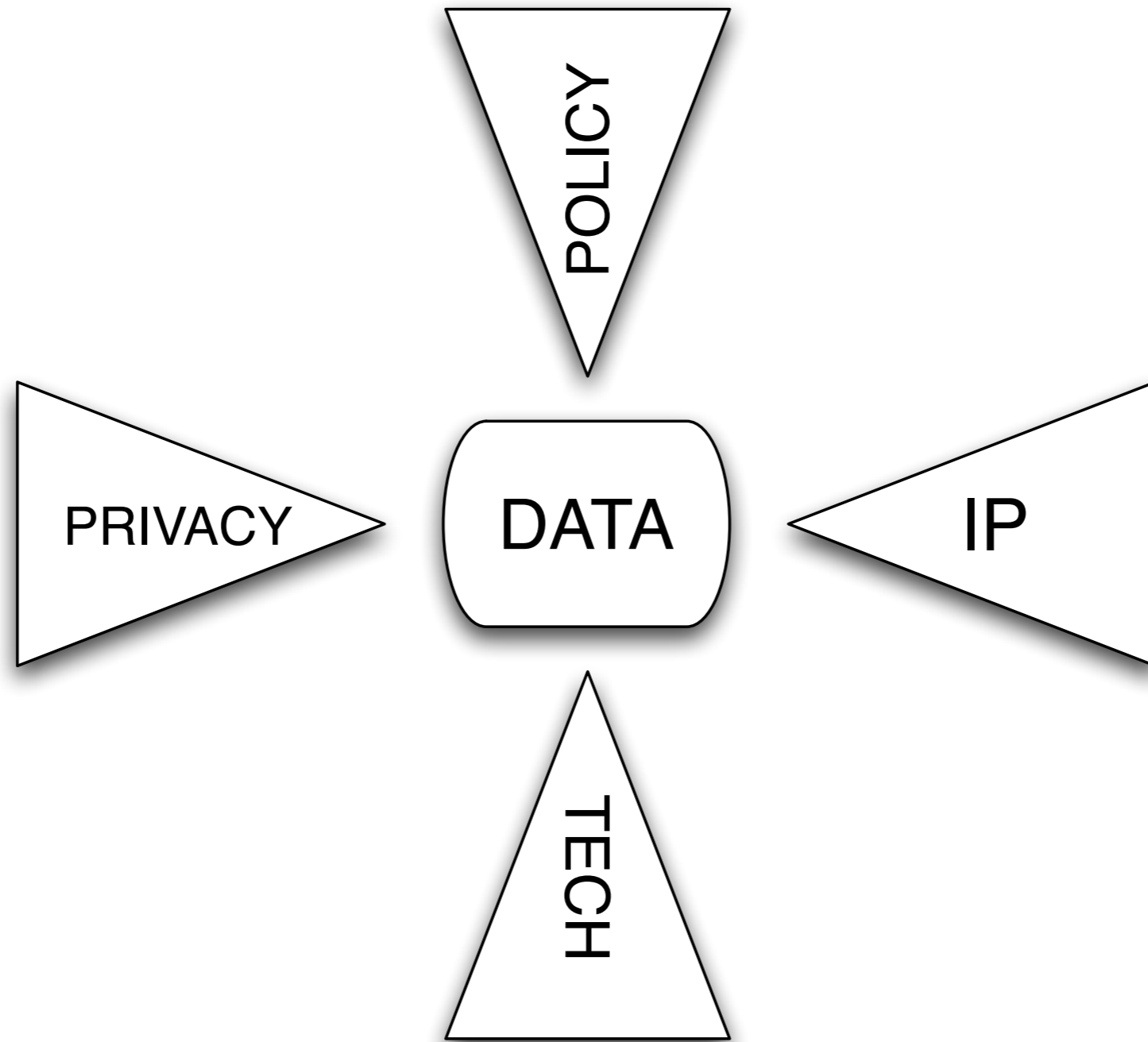
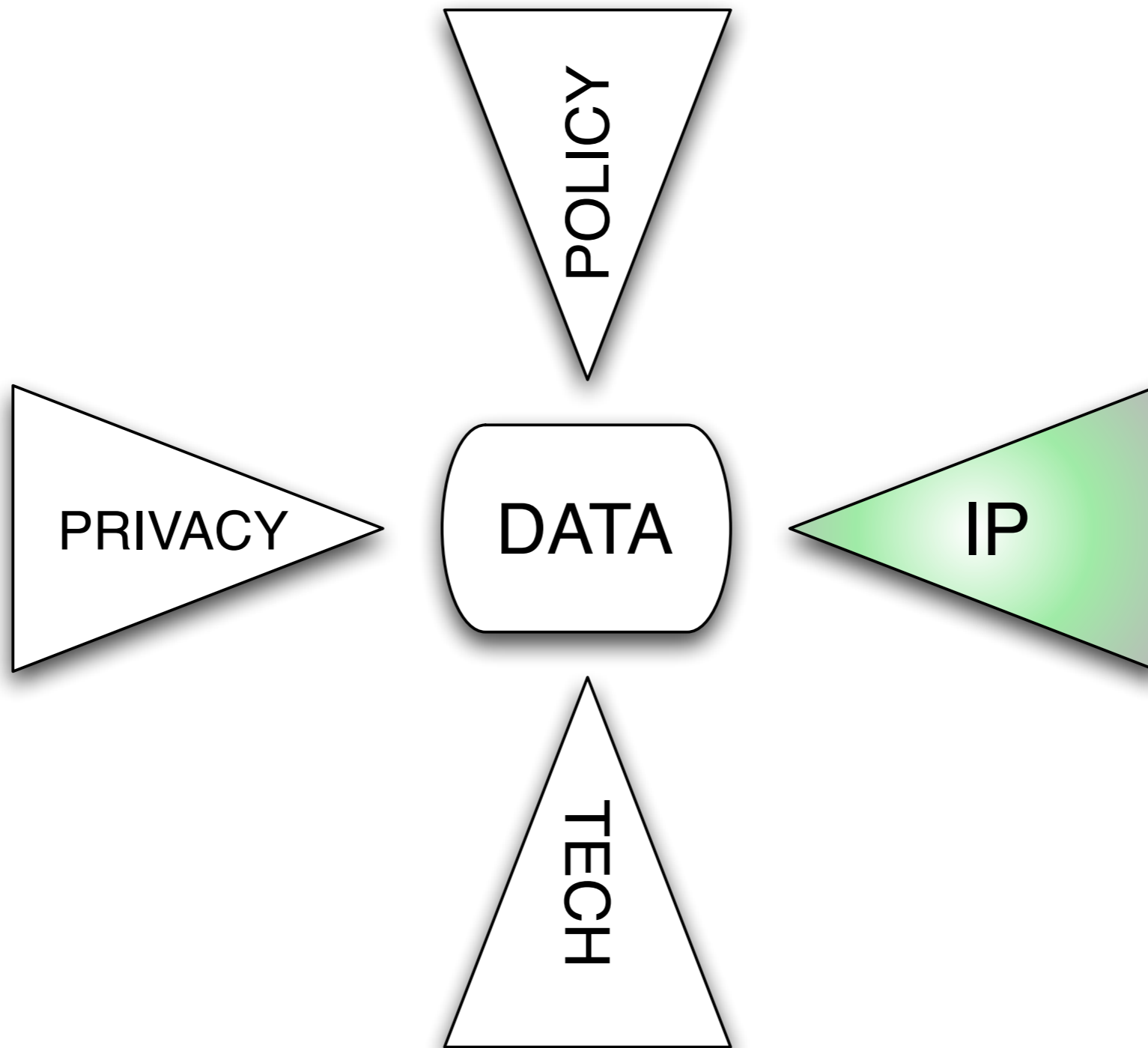# what is a regime that brings these together at *scale*?

pd or © or ©?

Database Rights

content agnostic

Schema
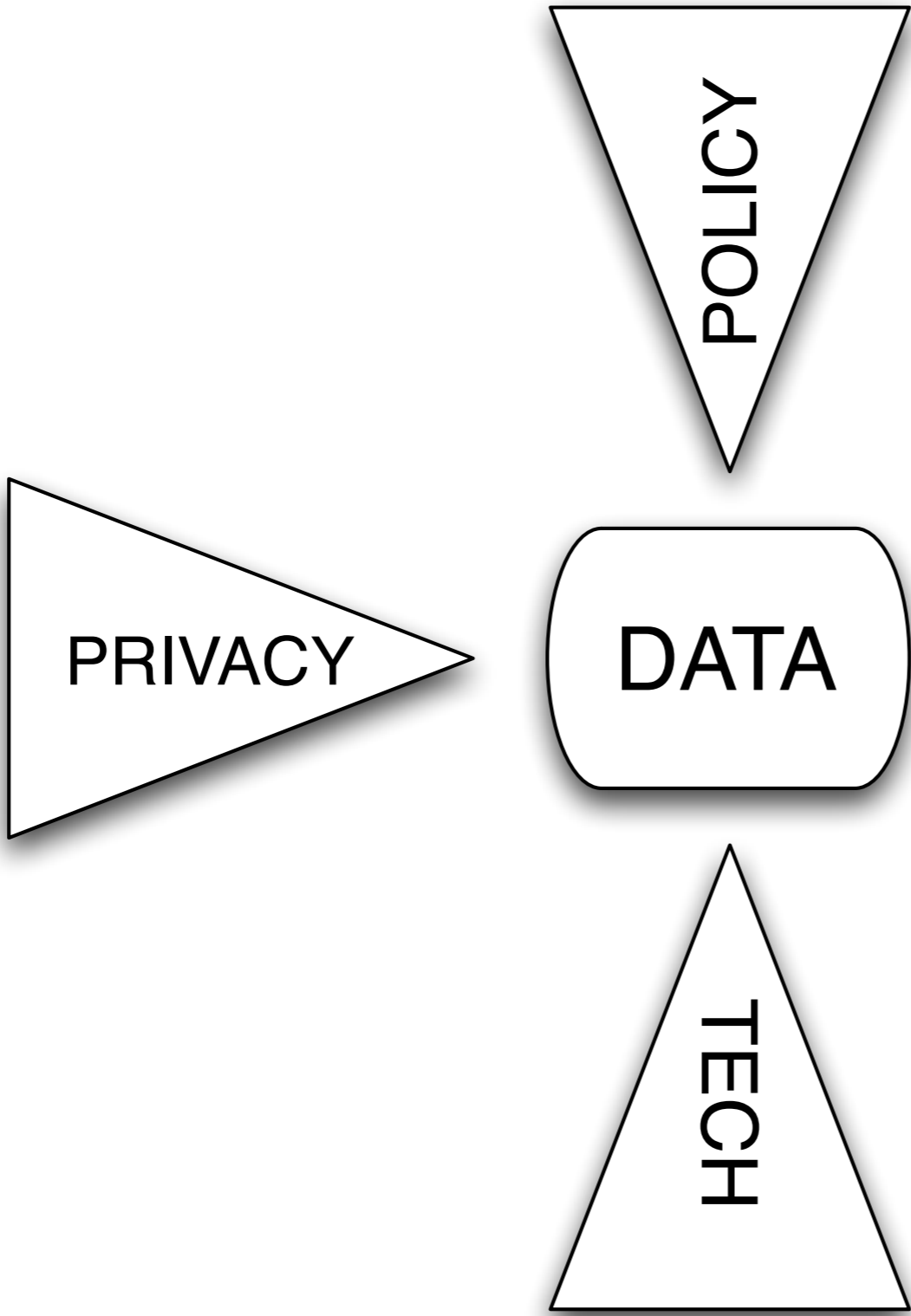model
design
diagram

© or ©?

Field Names

database

Database Rights

Tables

Data entry / output
sheets

© database software

Protocol for Implementing Open Access Data

2007

Status of this Memo

This memo provides information for the Internet community interested in distributing data or databases under an "open access" structure. There are several definitions of "open" and "open access" on the Internet, including the Open Knowledge Definition and the Budapest Declaration on Open Access; the protocol laid out herein is intended to conform to the Open Knowledge Definition and extend the ideas of the Budapest Declaration to data and databases.

This memo does not specify an Internet standard of any kind, but does specify the requirements for gaining and using the Science Commons Open Access Data Mark and metadata, by using legal tools and norms that conform to the protocol specified. This memo is available under the Creative Commons Attribution 3.0 (unported jurisdiction) license and will be submitted to the World Wide Web Consortium for consideration.

The terms MUST, MUST NOT, and SHOULD are used herein as defined in RFC 2119 ("Key words for use in RFCs to Indicate Requirement Levels").

1. Intellectual foundation for the protocol

The motivation behind this memorandum is interoperability of scientific data.

The volume of scientific data, and the interconnectedness of the systems under study, makes integration of data a necessity. For example, life scientists must integrate data from across biology and chemistry to comprehend disease and discover cures, and climate change scientists must integrate data from wildly diverse disciplines to understand our current state and predict the impact of new policies.

Panton Principles

Principles for Open Data in Science

Science is based on building on, reusing and openly criticising the published body of scientific knowledge.

For science to effectively function, and for society to reap the full benefits from scientific endeavours, it is crucial that science data be made **open**.

By open data in science we mean that it is freely available on the public internet permitting any user to download, copy, analyse, re-process, pass them to software or use them for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. **To this end data related to published science should be explicitly placed in the public domain.**

Formally, we recommend adopting and acting on the following principles:

1. Where data or collections of data are published it is critical that they be published with a clear and explicit statement of the wishes and expectations of the publishers with respect to re-use and re-purposing of individual data elements, the whole data collection, and subsets of the collection. This statement should be precise, irrevocable, and based on an appropriate and recognized legal statement

Web buttons

Get an **open data web button** for your project!

`OPEN DATA` `OPEN DATA`

Related Links

Open Knowledge Foundation - Working Group on Open Data in Science

Open Knowledge Definition

Is It Open Data?

Science Commons - Protocol for Implementing Open Access Data

2010

Monday, July 12, 2010

Deposited Set 1: 20th May 2010 - GSK TCAMS Dataset (hits from *P. falciparum* whole-cell screening)

Welcome to the Tres Cantos Antimalarial TCAMS dataset. Screening of approximately 2 million compounds in GlaxoSmithKline's screening library identified inhibitors of proliferation of *P. falciparum* strain 3D7 in human erythrocytes. The dataset contains the structures and screening data for over 13,500 compounds confirmed to inhibit parasite growth by more than 80% at 2 uM concentration. The compounds' activity against the multidrug resistant Dd2 strain has also been measured for comparison. In addition, we have included data for a human cell cytotoxicity selectivity screen and also deposited an indication of the 'promiscuity' of the hits (the IFI index) in other high-throughput assays at GSK. Finally, a potential mode of action and predicted *P. falciparum* targets are listed for selected compounds. All efforts have been made to ensure data quality and accuracy, but users are reminded that these data carry the usual caveats associated with results from large scale screening.

GSK does not guarantee the accuracy of any data, nor the suitability of the data for any purpose, in accordance with the EBI Terms of Use.

The chemical structures and the generated data are hereby made public under Creative Commons' CC0 license: http://creativecommons.org/publicdomain/zero/1.0/ as a resource for antimalarial lead identification and basic research into the druggable genome of *P. falciparum*.

GSK have committed to provide any corrections, additions and appropriate new annotations or data to ChEMBL-NTD

If you publish on, or wish to reference the GSK TCAMS set please include the link to ChEMBL-NTD (www.ebi.ac.uk/chemblntd) and adapt the following citation language: Francisco-Javier Gamo, Laura M. Sanz, Jaume Vidal, Cristina de Cozar, Emilio Alvarez, Jose-Luis Lavandera, Dana E. Vanderwall, Darren V. S. Green, Vinod Kumar, Samiul Hasan, James R. Brown, Catherine E. Peishoff, Lon R. Cardon and Jose F. Garcia-Bustos. Thousands of chemical starting points for antimalarial lead identification. *Nature* **465**(7296) 305-310 (2010) [pdf]

For further information please visit http://www.gsk.com/responsibility/access/rnd-neglected-tropical-diseases.htm

Sage Available and Transition Datasets

Sage Available Datasets are complete, globally-coherent datasets freely available from the Sage Commons Repository. The following table provides specifications of the datasets currently available and a PMID reference link or a text description. Researchers need only enter their name, organizational affiliation and email address on the download page. The repository packages include a readme file with descriptions and references and a (large!) compressed file with datasets and analyses. Researchers should send feedback on the program as well as questions, comments, and suggestions to repdata@sagebase.org.

Go to Repository Download Page

The **Sage Transition Dataset** table has the specifications of globally-coherent datasets that are in the process of being made publicly available from the Sage Commons repository. Interested researchers should check the Repository Download Page as partial datasets may be available. For more information on the status of the datasets contact repdata@sagebase.org.

Sage Available Datasets

Name	Tumor/ Tissue Type	Species	Disease	Approx. Num. Individuals	Investigator	Institution	Reference PMID/ Description
Mouse_CVD_Adipose_Liver_ Brain_Muscle_UCLA	Adipose, Liver, Brain, Muscle	Mouse	CVD	334	Jake Lusis	UCLA	description
Human_Cancer_HCC_HKU	HCC	Human	Cancer	250	John Luk	HKU	description
Human_CVD_Liver_Vanderbilt/ Pittsburg/StJudes	Liver	Human	CVD	517	Guengrich/Strom/ Schuetz	Vanderbilt Pittsburg	18462017

Data are the common wealth of humanity — *Adama Samassekou*
Convener of the UN World Summit on the Information Society

Overview of PIC

Who is Building PIC

Ethics and Norms of Data Sharing

Showcasing PIC

PIC in the Press

Welcome to
the Polar Information Commons (PIC):

Establishing the Framework for the Long-term Stewardship of Polar Data and Information

The polar regions are changing rapidly with dramatic global effect. Wise use of resources, astute management of our environment, improved decision support, and effective international cooperation on natural resource and geopolitical issues require a deeper understanding of, and an ability to predict change and its impact. Understanding and knowledge are built on data and information, yet polar information is scattered and scarce as well as temporally and spatially sporadic.

We are inspired by the Antarctic Treaty of 1959 that established the Antarctic as a global commons to generate greater scientific understanding. Correspondingly, we assert that data and information about the polar regions are themselves "public goods" that should be shared ethically and with minimal constraint.

ICSU
International Council for Science

CODATA

IASC

International Union
of Geodesy and Geophysics

Koninklijke Nederlandse
Akademie van Wetenschappen

SCAR
Scientific Committee on
Antarctic Research

science commons

World Meteorological Organization
Working together in weather, climate and water

we overestimate in the short term,
we underestimate in the long term.