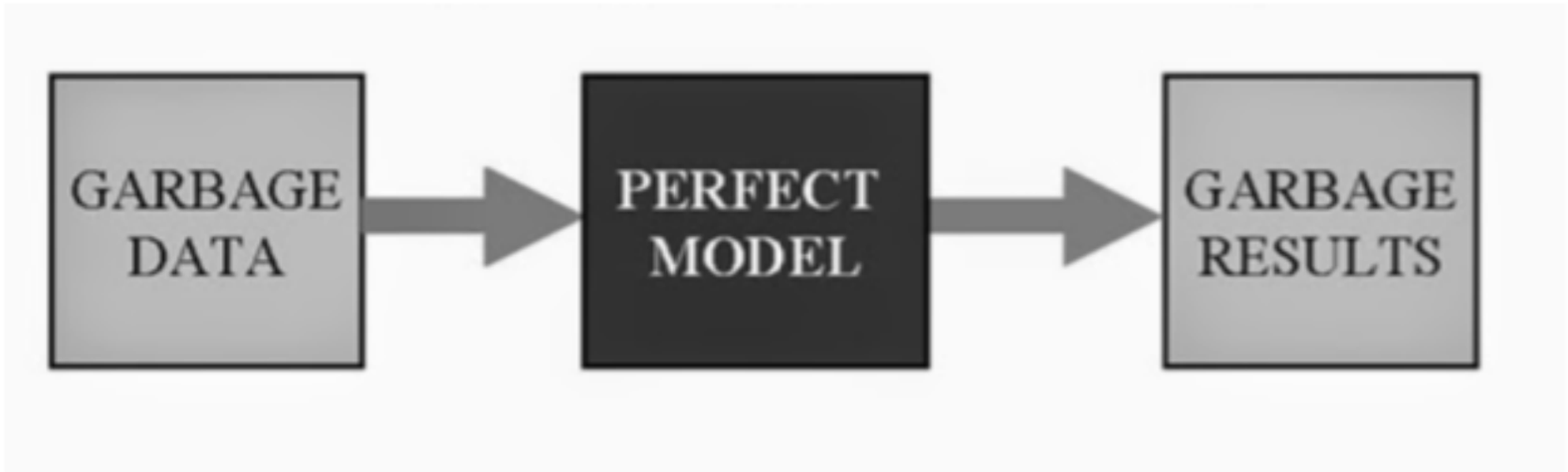
Summary + Presentations

Week 10 | INFX 551
Spring 2017

- **Summary Lecture**
    - Key concepts for INFX 551
    - What theory has to do with data curation
- **Presentations**
    - Groups presentations
    - Questions and Feedback
    - Interim critique (project swaps)
    - Final roadmap for deliverables

Data Curation

Old computer science saying "Garbage in = Garbage out"



Data curation says "Quality in = Quality out"

Data

(and in particular open data)

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Type vs Role
distinctions

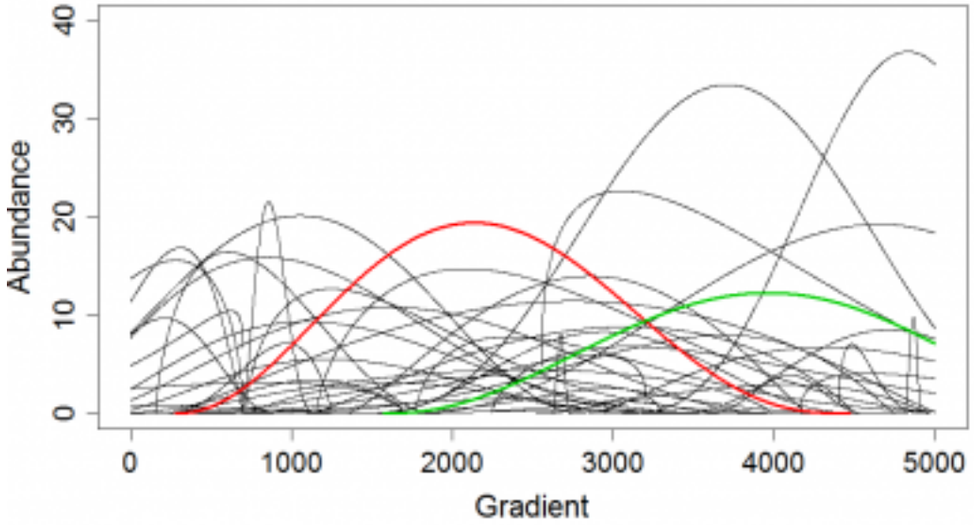
Type:
Donald Trump is a person.

Role:
Donald Trump is POTUS

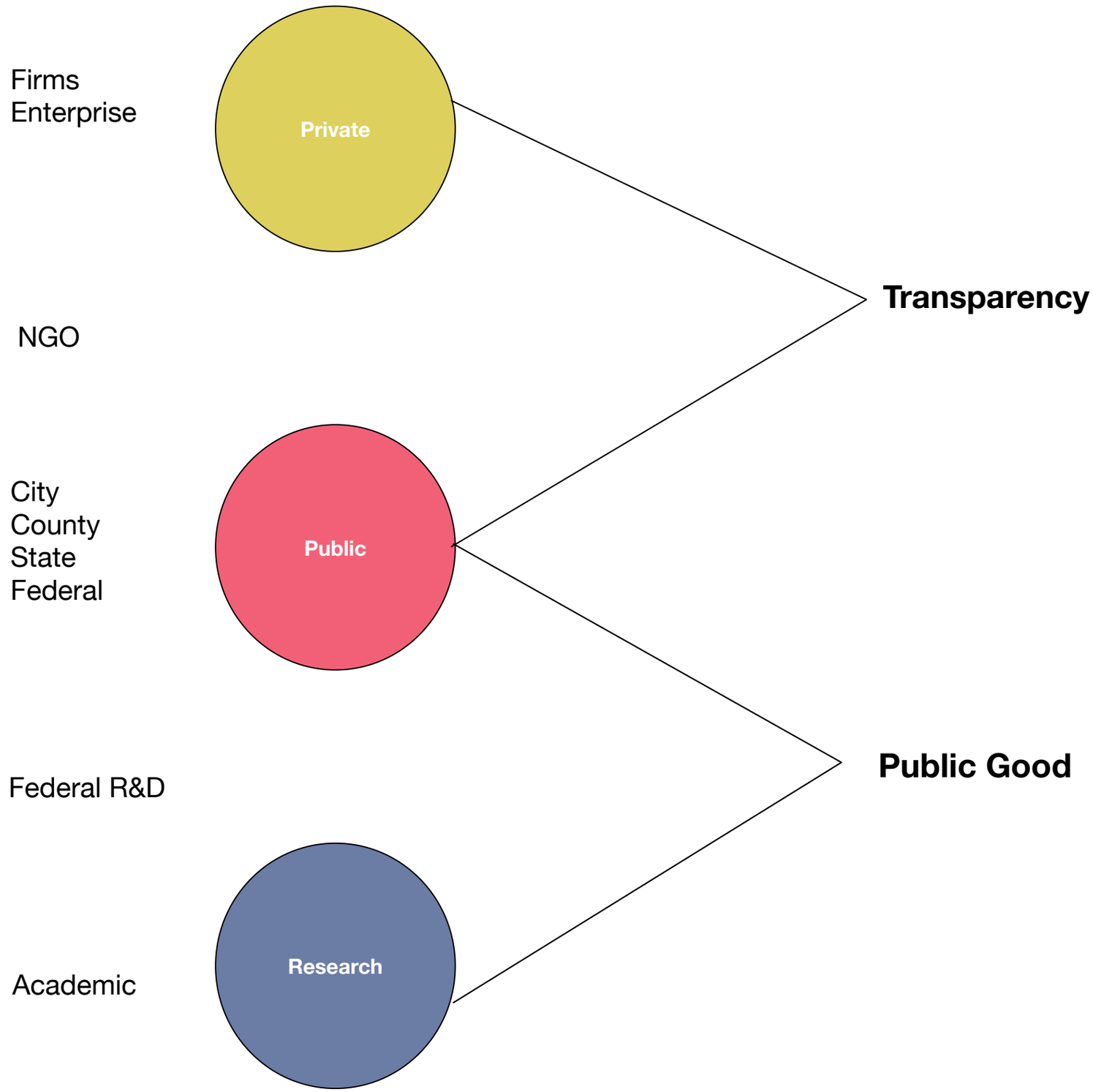
Type vs Role
distinctions

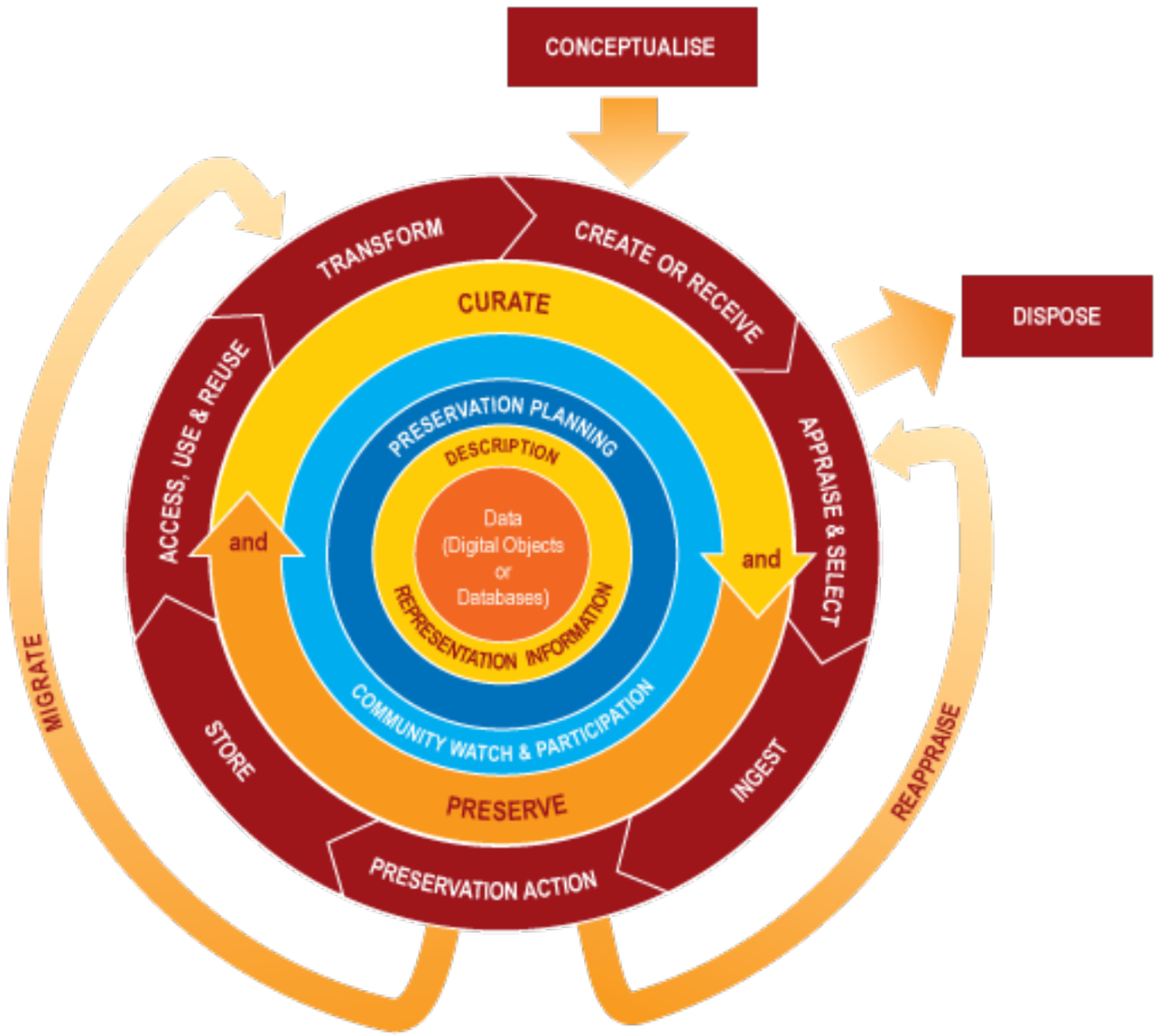
Type:
CSV

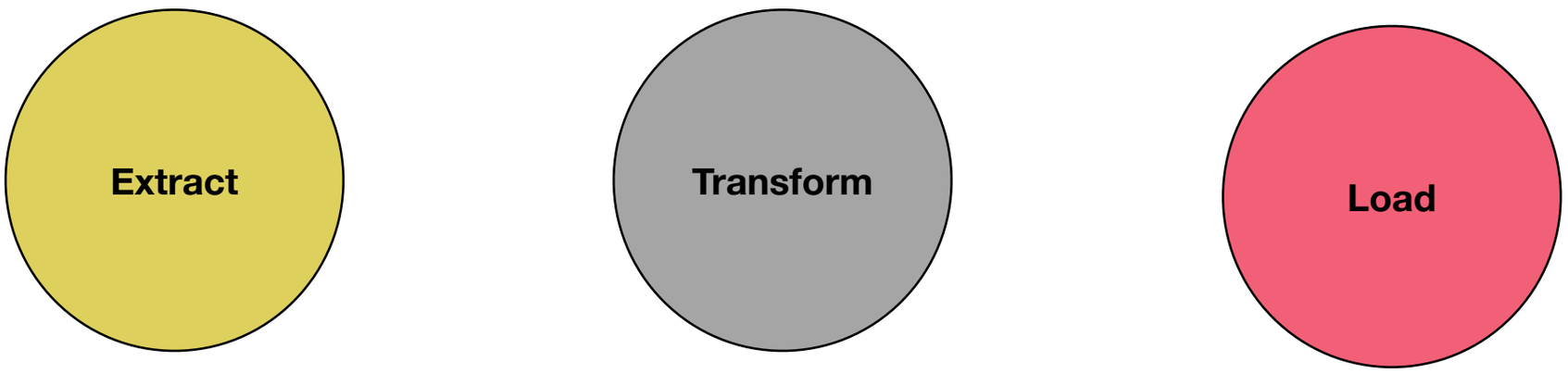
Role:
Evidence



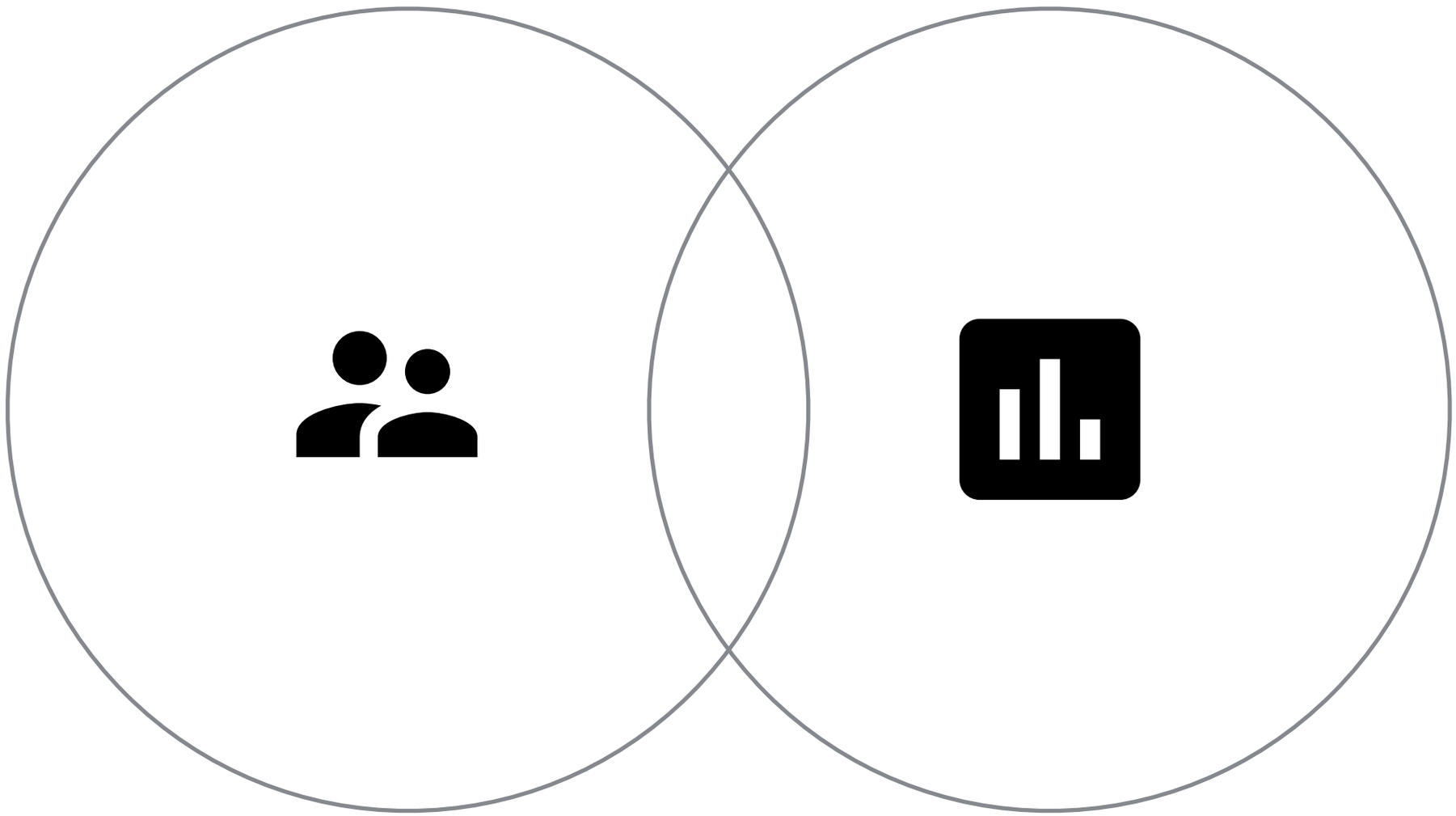
Curation


Firms
Enterprise
Private
NGO
City
County
State
Federal
Public
Federal R&D
Academic
Research
Transparency
Public Good


CONCEPTUALISE
DISPOSE
TRANSFORM
CREATE OR RECEIVE
CURATE
ACCESS, USE & REUSE
APPRAISE & SELECT
PRESERVATION PLANNING
DESCRIPTION
Data
(Digital Objects
or
Databases)
REPRESENTATION INFORMATION
COMMUNITY WATCH & PARTICIPATION
and
and
MIGRATE
STORE
PRESERVE
INGEST
REAPPRAISE
PRESERVATION ACTION


Extract
Transform
Load

Repositories



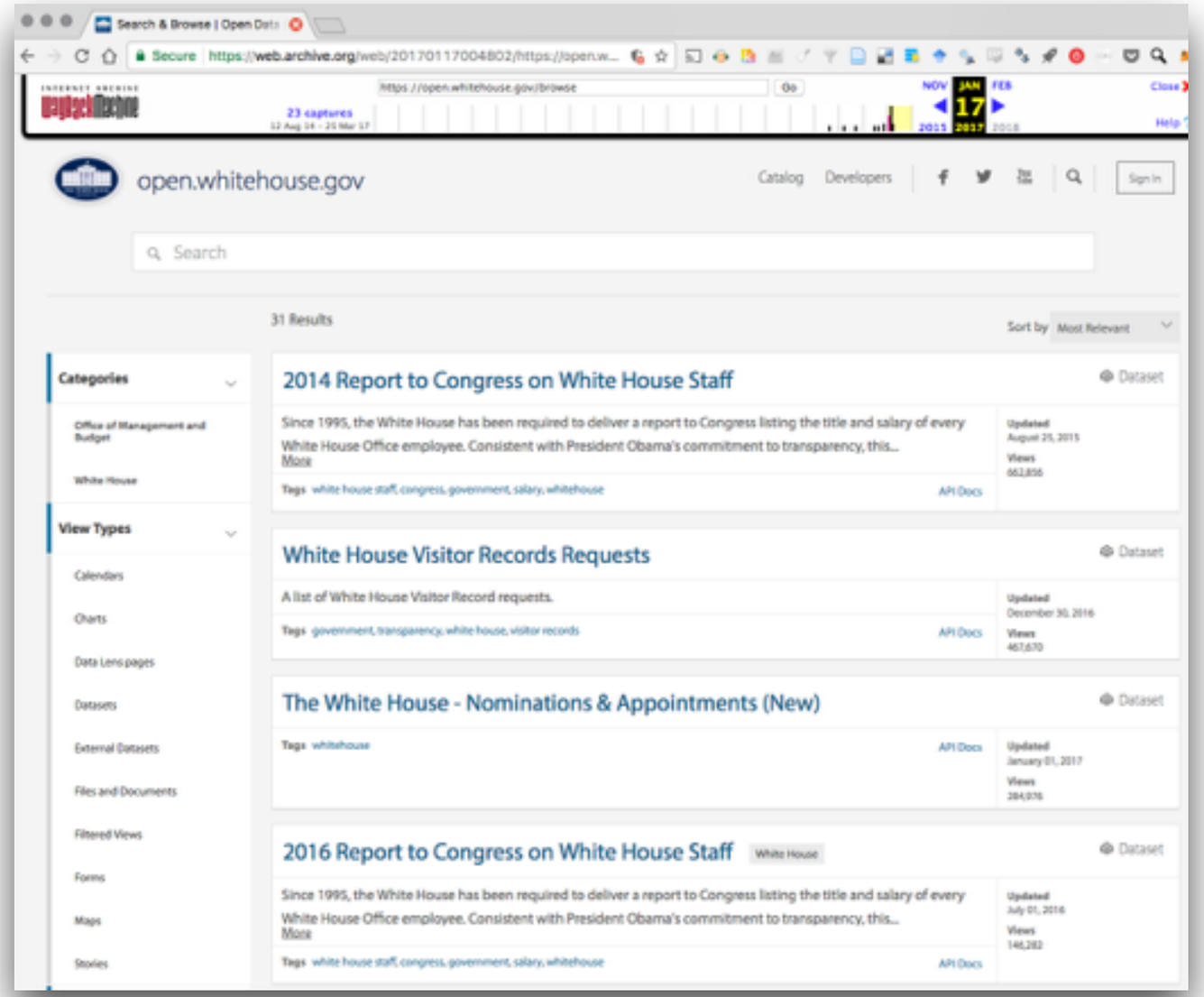
Data repositories are **sociotechnical** infrastructures.

Mutual constitution of people and technologies, as they are embedded in political, economic, and ecological systems of collective action.

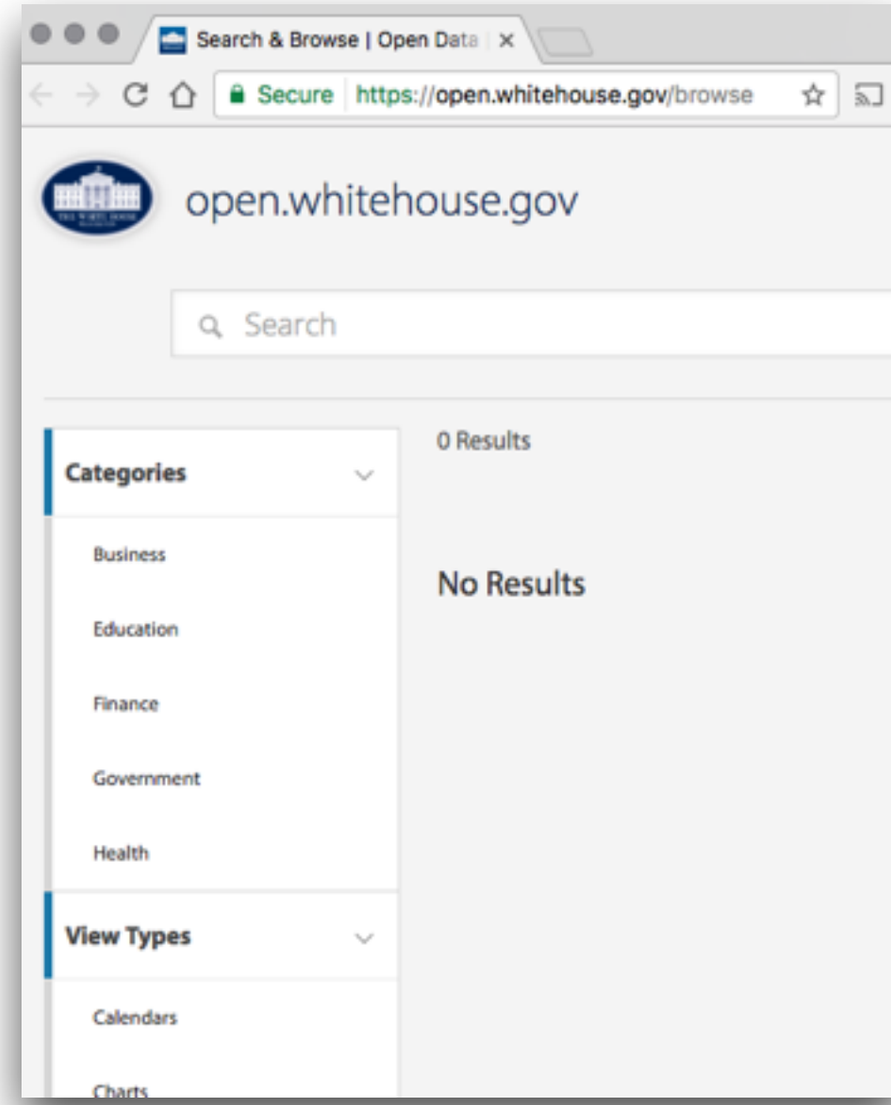
Data repositories are sociotechnical **infrastructures**.

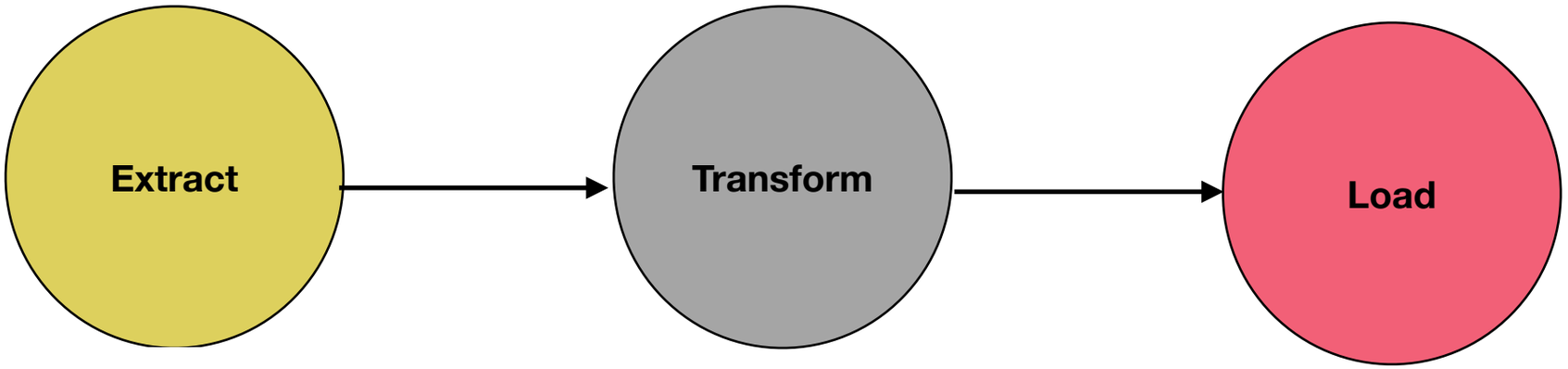
… relational
… embedded
… learned (as opposed to sensical)
… made up of standards
… built on other infrastructures
… often invisible (until break down)

January 17th



Current




Extract
Transform
Load

User-centered metadata

Functional Requirements —> Technical Specifications

Systems Development for Data Curation

User Stories —-> Metadata Schemas

#1 Searching for data about a subject.

In order to inform my decision about whether to expand my garden to begin selling produce, as a casual gardener, I would like to find data about volume of sales at farmers' markets for various kinds of produce in my region.

How to satisfy? Provide metadata or search functionality about which localities are included in a given dataset. Include easy to access variable lists. Provide mapping functionality.

User Stories —-> Metadata Schemas

Will an existing schema meet all of these requirements?

We determined that DCAT would provide an appropriate metadata standard by which we can meet virtually all of our requirements. DCAT is widely used in the data sectors relevant to recycling data and numerous extensions and modifications of the standard exist for public sector data, suggesting that it can be adapted to new needs and ideas. In addition, numerous examples and guides for use of DCAT exist online – a list of potentially useful DCAT resources we discovered during this activity and may wish to consult in future is included herein as Appendix A for our reference.

We anticipate that our list of required metadata elements could be mapped to DCAT as follows:

Proposed Element List	Corresponding DCAT Property/ies	Preliminary Notes for Development of Full Application Profile
Unique identifier	dct:identifier	mandatory; define identifier scheme
Title	dct:title	mandatory
Description	dct:description	mandatory
Author and/or publisher	dct:creator; dct:publisher	mandatory; unbounded
Contact point/responsible party	dcat:contactPoint	optional; unbounded

The Halting Problem of Knowledge Representation

Expressivity VS. Tractability

The more expressive we make a knowledge representation system, the less tractable it is in terms of generating, managing, and computing against the KR.

What makes a good schema?

*Avoids halting problem
*Is expressive and meaningful to your domain
*Modular architecture

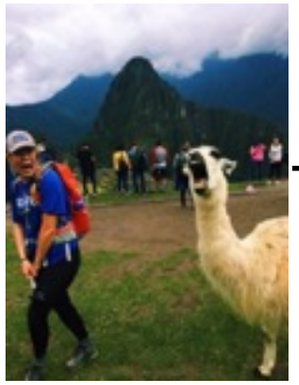
Preservation

Digital objects are living in a material world.
(they are stored on media that eventually break down)

Preservation

```
0008810: 2828 2828 2828 2828 2828 2828 2828 28ff  ((((((((((((((.
0008820: c000 1108 0379 024f 0301 1100 0211 0103  .....y.O........
0008830: 1101 ffc4 001c 0000 0203 0101 0101 0000  ................
0008840: 0000 0000 0000 0003 0401 0205 0600 0708  ................
0008850: ffc4 004a 1000 0202 0103 0303 0203 0604  ...J............
0008860: 0306 0402 0b01 0203 1100 0412 2105 3141  ............!.1A
0008870: 1322 5106 6132 7181 0714 2342 91a1 52b1  ."Q.a2q...#B..R.
0008880: c1d1 1533 f024 3462 72e1 f108 1643 8235  ...3.$4br....C.5
0008890: 5392 b217 2573 a2c2 d236 4463 ffc4 001a  S...%s...6Dc....
00088a0: 0100 0301 0101 0100 0000 0000 0000 0000  ................
00088b0: 0000 0102 0304 0506 ffc4 0033 1100 0202  ...........3....
00088c0: 0202 0104 0201 0304 0301 0002 0300 0102  ................
00088d0: 1103 2112 3141 0413 2251 3261 7105 1442  ..!.1A.."Q2aq..B
00088e0: 2333 8191 15a1 b152 62f0 c1d1 e1ff da00  #3.....Rb.......
00088f0: 0c03 0100 0211 0311 003f 00dd 9b4e dae8  .........?...N..
0008900: 77f4 b8cb c2b2 9500 c772 283c 8b3f ea33  w........r(<.?.3
```

AlpacaScream.jpg



Is this a Jpeg?

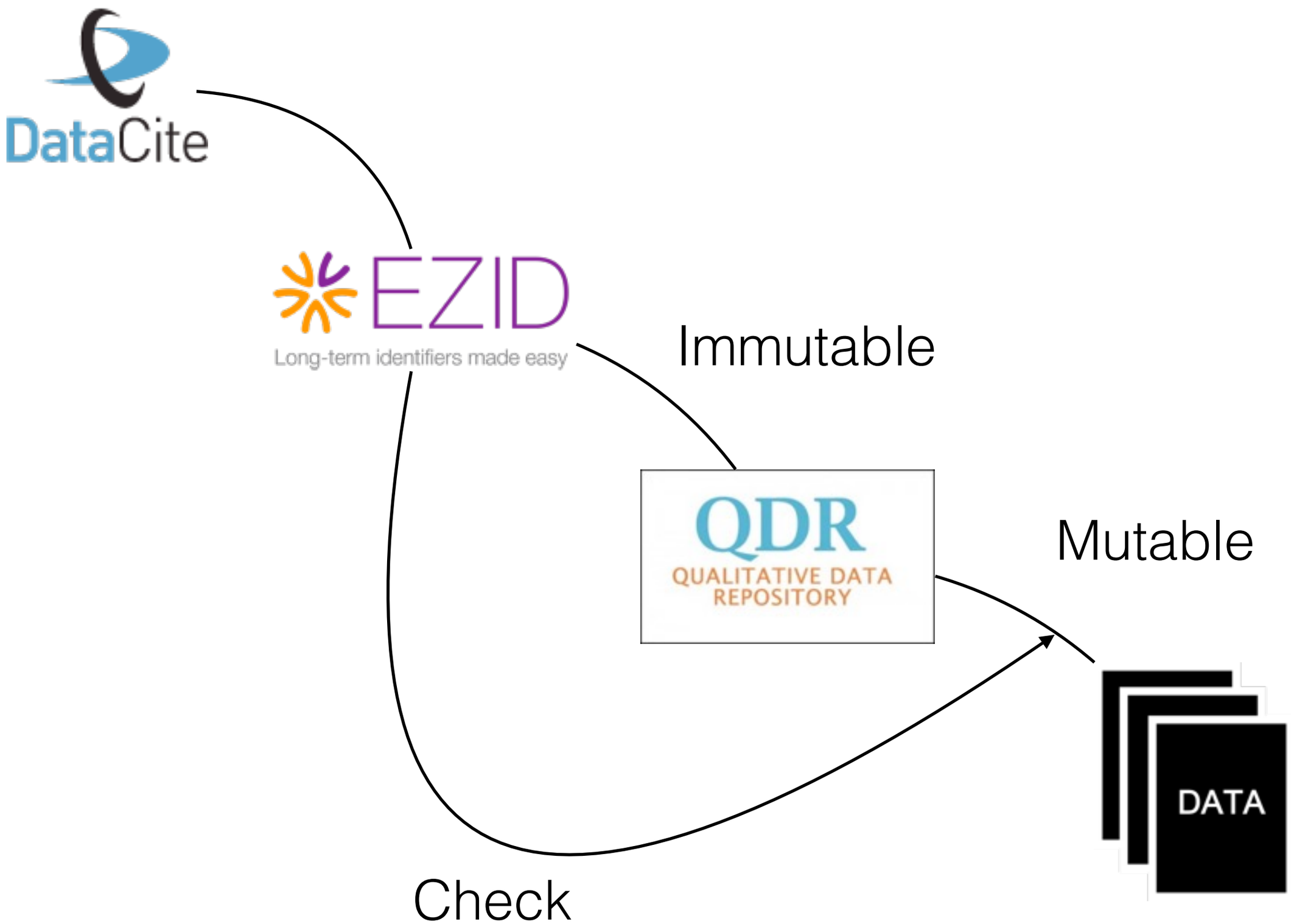
Yes

Data Identity on the Web

ID’s are signposts to resources on the web– they need to point to a thing (or give users a clear idea of where to find the thing once they get there)

Just as an address won’t prevent a house from being demolished or remodeled, an ID will not prevent “content drift” or data from being deleted

They don’t make content more reliable, citable, or trustworthy - they are social cues.


DataCite
EZID
Long-term identifiers made easy
Immutable
QDR
QUALITATIVE DATA REPOSITORY
Mutable
DATA
Check

Policy

Policy background for open data in the USA…

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

THE DIRECTOR

May 9, 2013

M-13-13

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Sylvia M. Burwell
Director

Steven VanRoekel
Federal Chief Information Officer

Todd Park
U.S. Chief Technology Officer

Dominic J. Mancini
Acting Administrator, Office of Information and Regulatory Affairs

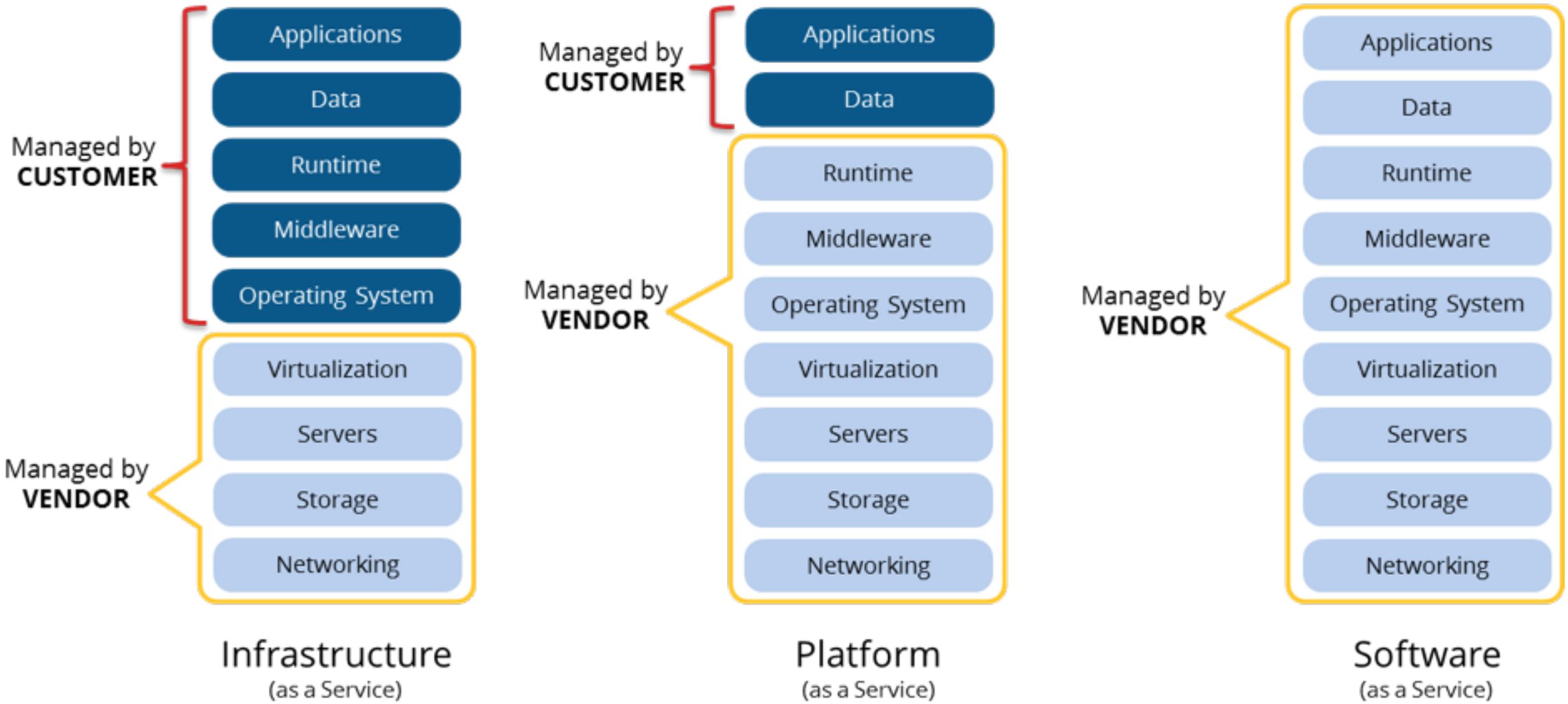
SUBJECT: Open Data Policy—Managing Information as an Asset

"…this Memorandum requires agencies to collect or create information in a way that supports downstream information processing and dissemination activities."

Models of sustainability & services

Funding Models	Potential for Economic Stability Needed for Long-Term Sustainability	Potential for Open Access to Research Data	Potential for Equity for Deposits by Individual Researchers	Potential for Equity for Universities/ Institutions
Membership Dues	Moderate; subject to institutional budgets and priorities	Low	Moderate	Low
Submission Fees	Low to Moderate; subject to policies of funding agencies and publications;	High	Low; costs transferred from end users to data producers	Low
Institutional support	Moderate; subject to institutional budgets and priorities	High	Low	Low
Federally-sponsored Special Projects	Low; subject to changes in national research priori-	High	Limited to designated research	High

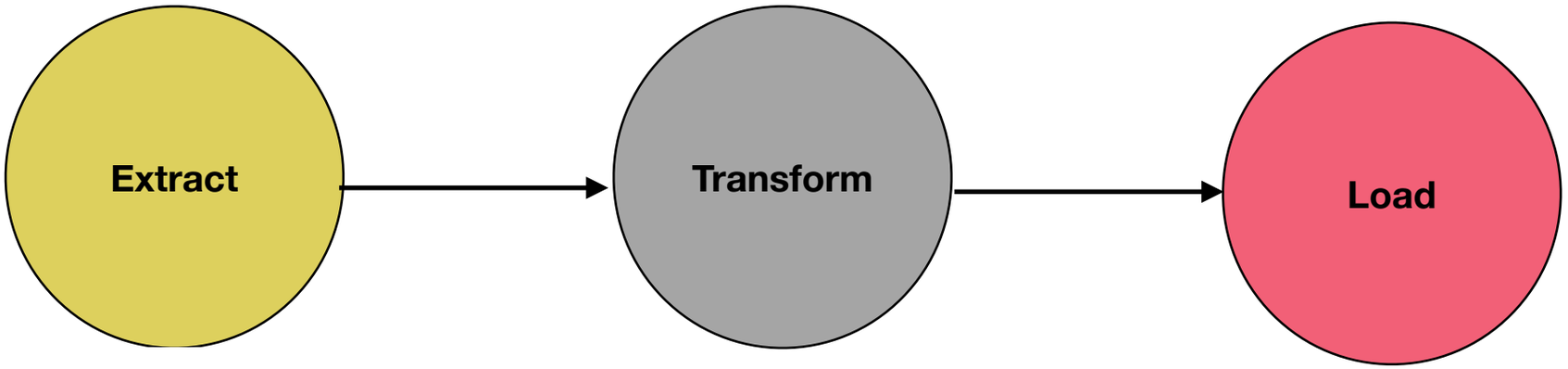
http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf



amazon web services

heroku

Google Apps


Extract
Transform
Load

"very impressive, professor...,
but does it
work in theory?
F
AUGUT

What do I mean by ‘theory’ in this context ?

…a conceptual tool for making sense of a domain…

- **Descriptive Power**: provide a conceptual frame-work that helps us make sense of and describe the world

- **Rhetorical Power**: help us talk about the world by naming important aspects of the conceptual structure and how it maps to the real world.

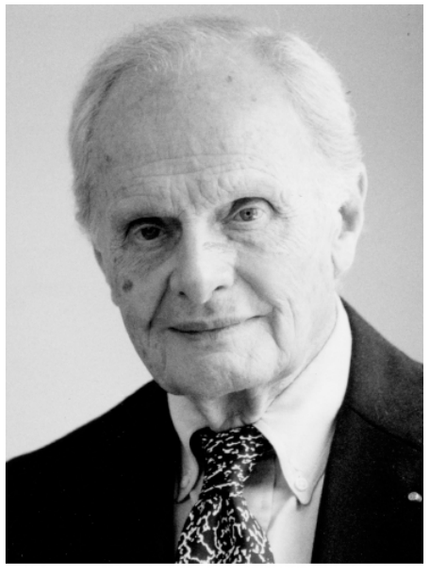
- **Inferential Power**: help us predict outcomes (or reliably reason about them), but also tell us ‘where to look’ when we’re studying new phenomena

- **Applicability**: be able to translate to the real world on pragmatic terms (that is - it should help us bridge a gap between description/ prediction and action)

Straw man:

Does technology shape our actions, or do we shape technology to perform certain actions?

“Technological Momentum”



Thomas P. Hughes

….technologies are designed to solve a problem first … early in their existence they are often malleable to demands of users, and to design adjustments….as they become more broadly adopted and used, they being to lock users and designers in to certain modes of action…

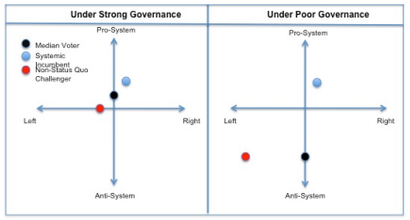
Descriptive Power
Rhetorical Power
Inferential Power

QDR BETA

QUALITATIVE DATA REPOSITORY

FIGURE 3

POLITICIZING GOVERNANCE ENABLES LEFT-RIGHT POLARIZATION



As parties are not unitary actors, the effects of governance on the strategic landscape manifest themselves in the factional contestation that occurs within non–status quo parties and blocs. High-quality governance tends to empower moderate factions within these parties and blocs, which can argue more convincingly for the necessity of left-right programmatic moderation for electoral viability. Meanwhile, radical factions tend to have great difficulty articulating an alternative path to winning power. How do they propose to actually win an electoral majority if their left-right program diverges so greatly from popular preferences and there is no other dimension of contestation on which they might capitalize? Further, in multiparty systems, moderates may play another trump card—the strategic necessity of forming coalitions, and consequently the pragmatic necessity of moderating their program so that it will be palatable to coalition partners—to which radicals often have little response. Low-quality governance, in contrast, tends to undercut moderates and empower more radical factions within parties and blocs. Moderate arguments about the necessity of programmatic moderation ring hollow when radicals can articulate an alternative path to power that emphasizes

governance-based appeals and critiques of the track records of incumbents in this respect (as heuristically shown in the right panel of Figure 3). Further, moderate arguments about coalition building can now be turned on their head, as radicals can convincingly argue that joining coalitions with status quo parties is undesirable, since it jeopardizes the distinctive identity unsullied by association with status quo parties, undercutting the seemingly inescapable logic of coalition building.

While these theoretical propositions pertain in the abstract to parties of both left and right, their impact on the left is particularly relevant for understanding and explaining variation in the contemporary Latin American case universe. As previously noted, instances of high polarization in contemporary Latin American have been marked by the presence of radicalized left parties, which have challenged neoliberal orthodoxy and in at least a subset of cases have framed their projects in uncompromising terms.[8] I therefore specifically argue that governance levels during the pre–Left Turn period explain subsequent patterns of polarization via their decisive impact on factional conflict within the left. Where governance was relatively strong, moderate factions within major left parties triumphed over radicals, establishing strong social democratic parties that would anchor centripetal party systems. Where governance was poor, radical factions within the left were able to buck moderates, leading to either established left party challengers consolidating in a radical direction under their leadership or these parties breaking down and the subsequent construction of new parties or blocs that fused old and new radical actors.

While the focus of the argument is to explain variation in the level of polarization in party systems, another concern is to further explain whether cases of high polarization took on highly contentious or more pragmatic patterns of contestation. In this respect, the key variable was whether the radical left forces that eventually emerged took the form of established challengers, as in Nicaragua and El Salvador, or the form of coalitions led by outsider forces, as in Venezuela, Bolivia, and Ecuador. While in each pattern radicals were strengthened in their internal struggles with moderates by the poor quality of

[8] In the case of El Salvador, radicalism on the right has also certainly contributed to the overall level of ideological differentiation within party systems. Nevertheless, the more constant pattern involves radicalized parties on the left.

A BROAD ASSESSMENT ACROSS LATIN AMERICA

An initial empirical evaluation broadly examines how well the two key independent variables predict the outcome of a highly polarized party system developing in the 2000s. The first variable is the presence of electorally relevant parties of the left and the right in the pre–Left Turn period, with particular attention to the presence of parties of the left (the source of divergence across cases). The number of countries lacking electorally relevant leftist parties is noteworthy. Many cases possessed party systems dominated by traditional parties into which the left historically struggled to make substantial inroads, such that no electorally relevant left parties existed in the immediate pre–Left Turn period (Panama, Honduras, Guatemala, Paraguay, Costa Rica, Colombia).[9] Others possessed party systems historically marked by strong labor-based populist parties, which had crowded out and marginalized the left, contributing to a similar absence of electorally relevant left parties (Argentina, Peru).[10]

Perhaps unsurprisingly, we do not observe high levels of party-system polarization in any of these countries during the Left Turn period. In some of the cases, polarization has continued to be attenuated by the lack of supply on the left side of the ideological spectrum. In other cases, politicians without genuine roots in leftist parties have claimed the mantle of the left, often for seemingly strategic purposes, but not in ways that introduced truly high levels of polarization into party systems. In Argentina, Néstor and Cristina Kirchner have moved the Peronist Party vaguely back to the left side of the ideological spectrum and governed in ways that have inspired significant controversy. Nevertheless, the lack of clear ideological definition to the Peronist Movement and the reticence of the Kirchners to fully embrace the scorched-earth style of

[9] In Colombia, the AD/M-19 (Democratic Alliance/19th of April Movement, Alianza Democrática/ Movimiento de 19 de Abril), a fledgling party formed by the demobilized M-19 guerrilla movement, won a surprisingly high share of the vote in the 1990 congressional and then about 8.5 percent in the new congressional elections of 1991. The party quickly collapsed and was electorally irrelevant by the 1994 elections, however, such that it does not meet the paper's standards for electoral relevance.

[10] In Peru, the leftist IU (United Left, Izquierda Unida) had won a surprising 27 percent of the lower house seats in the 1985 congressional elections. However, IU's seat share dwindled to 9 percent in the 1990 elections and the party became electorally irrelevant soon afterward. Obviously, the decline of IU in particular and absence of left parties in general was likely related to the regime transition under Alberto Fujimori. For present purposes, the important fact is simply that an electorally relevant left ceased to exist after about 1990.

“conceptual tool for making sense of a domain"

- helps to understand the preferences of a very traditional domain of scholarship (political science) (form determines practice)
- guides action - > we need to build annotation tools (instead of dynamic documents)
- (not Hughes’ intention, but…) technological momentum - solve social problem first, and lock users in to new mode of working (over time)

(Socialist Convergence) faction.[14] Articulação leaders such as Lula and José Dirceu justified these strategic decisions through appeals to electoral viability, arguing that a more pragmatic politics was necessary if the PT wanted to win power.[15]

In the context of extremely poor governance during the early 1990s, the PT's moderate leadership was met by an unprecedented challenge to its leadership (Hunter 2010). Radical factions launched a series of broadsides in the partisan press and other PT fora, lambasting the party's leadership for seeking to make the PT part of the Brazilian political status quo rather than a challenger to that established order. Fueling these critiques were consistent references to the context of poor governance and the opportunities it afforded the party. Radicals blasted the leadership for "not providing a single concrete response to the crisis of governance" in the country and argued that "in some moments popular dissatisfaction explodes, looking for new alternatives."[16] In this context, Articulação was dealt a massive blow with the defection of a substantial part of its leadership and activist base, which formed the new faction "Articulação de Esquerda" and joined in an alliance in advance of the 1993 Party Congress with longer-standing radical factions such as Na Luta PT and Força Socialista.

The 1993 party congress saw radicals win the leadership of the PT for the first time in its existence, a swing in which the defection of Articulação de Esquerda was decisive. The arguments fueling the radical uprising, as seen in Articulação de Esquerda's manifesto prepared for the congress, directly played on the governance crisis and the strategic opportunities it presented. The faction argued that the PT's most viable path to victory in 1994 involved clearly distinguishing themselves from systemic parties implicated in poor governance, that "defeats in recent elections show clearly that the PT loses when it adopts ambiguous behavior," and that "it is unacceptable that, on account of poorly thought out electoral calculations, our party fails to present a comprehensive alternative to the Brazilian crisis."[17]

---

[14] On these points, see *Boletim Nacional* (1992.2, 1992.5a) and *Brasil Agora* (1993.5).
[15] For an example of this kind of viability appeal from the moderate leadership, see Boletim Nacional (1992.12).
[16] See *Boletim Nacional* (1992.5b, 1993.3a). Translations, here and throughout paper, are by the author.
[17] *Boletim Nacional* (1993.3b).

Public user:qdr@hypothes.is Sign up / Log in

1992.2

Annotation for Transparent Inquiry (ATI) QDR

Full Citation: Boletim Nacional. 1992.2. "DN aprova tática e política de alianças." Collected in Partido dos Trabalhadores Historical Archives, Projeto Memória e História, Fundação Perseu Abramo. Leiden, Netherlands: IDC.

Click to access full source.

Analytic Note: This source demonstrates the PTs new more pragmatic direction in the early 1990s in the controversial area of alliance policy, with the partys national directorate approving the adoption of a formal alliance policy that sanctioned electoral pacts with many different parties to the left of center, including even center-left parties such as the PSDB among these potential allies and using terminology such as flexible to characterize the new direction.

Excerpt: Embora seja necessária uma política de alianças nacional, as eleições de 92 exigem que se leve em conta a diversidade de realidade locais. Nossa política deve ser flexível: é uma política de alianças nacional, mas que se articula também com a dinâmica das disputas municipais. O PT tem desenvolvido alianças a partir de um projeto de construção de uma alternativa democrática e popular. Neste campo, se encontram os PCs, o PSB, o PV, o PSDB e o PPS. Ao propor uma política de alianças com o PSB, com os PCs e com o PPS, não desconhecemos as contradições e diferenças políticas que mantemos com esses partidos, mas levamos em consideração, principalmente, a posição política na luta contra o governo Collor, as nossas alianças em 89 na Frente Brasil Popular, o apoio de Roberto Freiré e do PSB a LULA no 2º turno, o amplo arco de alianças e acordos que realizamos em 89 e 90, inclusive governando algumas cidades conjuntamente.

Translation: Although a policy of national alliances is necessary, the elections of 92 require taking into account the diversity of local realities. Our policy should be flexible: it is a policy of national alliances, but it also articulates with the dynamics of local disputes. The PT has developed alliances for a project to build a democratic and popular alternative. In this group, we find the PCs, PSB, PV, the PSDB and the PPS. By proposing a policy of alliances with the PSB, with PCs and with PPS, we are not unaware of contradictions and political differences we have with these parties, but we take into account mainly the political position in the fight against Collor government, our 1989 alliances in the Brazil Popular front, the support of Roberto Freire and the PSB to LULA in the 2nd round, the wide arc of alliances and agreements that we made in 89 and 90, including jointly ruling some cities.

Less

- Theory as a conceptual tool for making sense of a domain.
    - Has certain powers, which helps us explain curation problems to other people, and to ourselves
    - Those powers are descriptive, rhetorical, inferential, and applicability
    - Using theory to guide curation actions can bring clarity to certain tasks, and
    - Help us design services and tools that can shape and be shaped by user needs.