

La privadesa en l'estadística oficial: el secret estadístic i el dret a la informació

Enric Ripoll i Font

“La lluita contra el Big brother: tecnologia, percepció social i legislació sobre privadesa en la societat de la informació”

(URV, 25 de juny de 2007)

Privadesa en l'estadística oficial (guió)

1. Marc legal en protecció de dades personals/secret estadístic

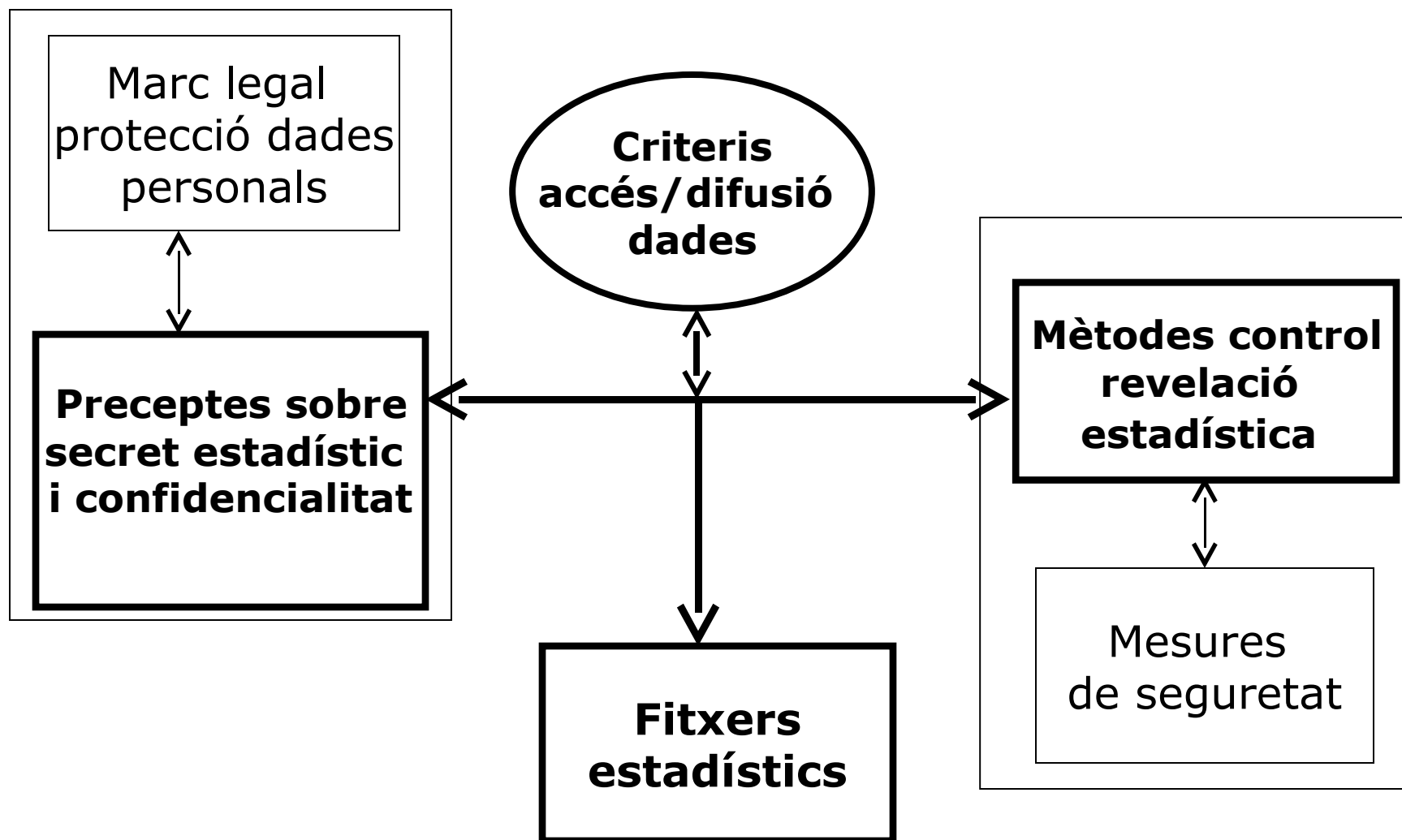
- 1.1. La interacció entre la LOPD i la LEC: aplicació i diferències
- 1.2. Marc legal i preceptes sobre la preservació del secret estadístic

2. Control de la revelació estadística: tecnologia

3. Criteris sobre preceptes i mètodes

- 3.1. Opcions i reptes en l'estadística oficial: el cas de l'Idescat
- 3.2. Criteris corporatius d'accés i difusió de microdades: el cas d'Idescat
- 3.3. El futur Registre de Fitxers Estadístics del Sistema estadístic català

Elements clau en les polítiques de confidencialitat estadística



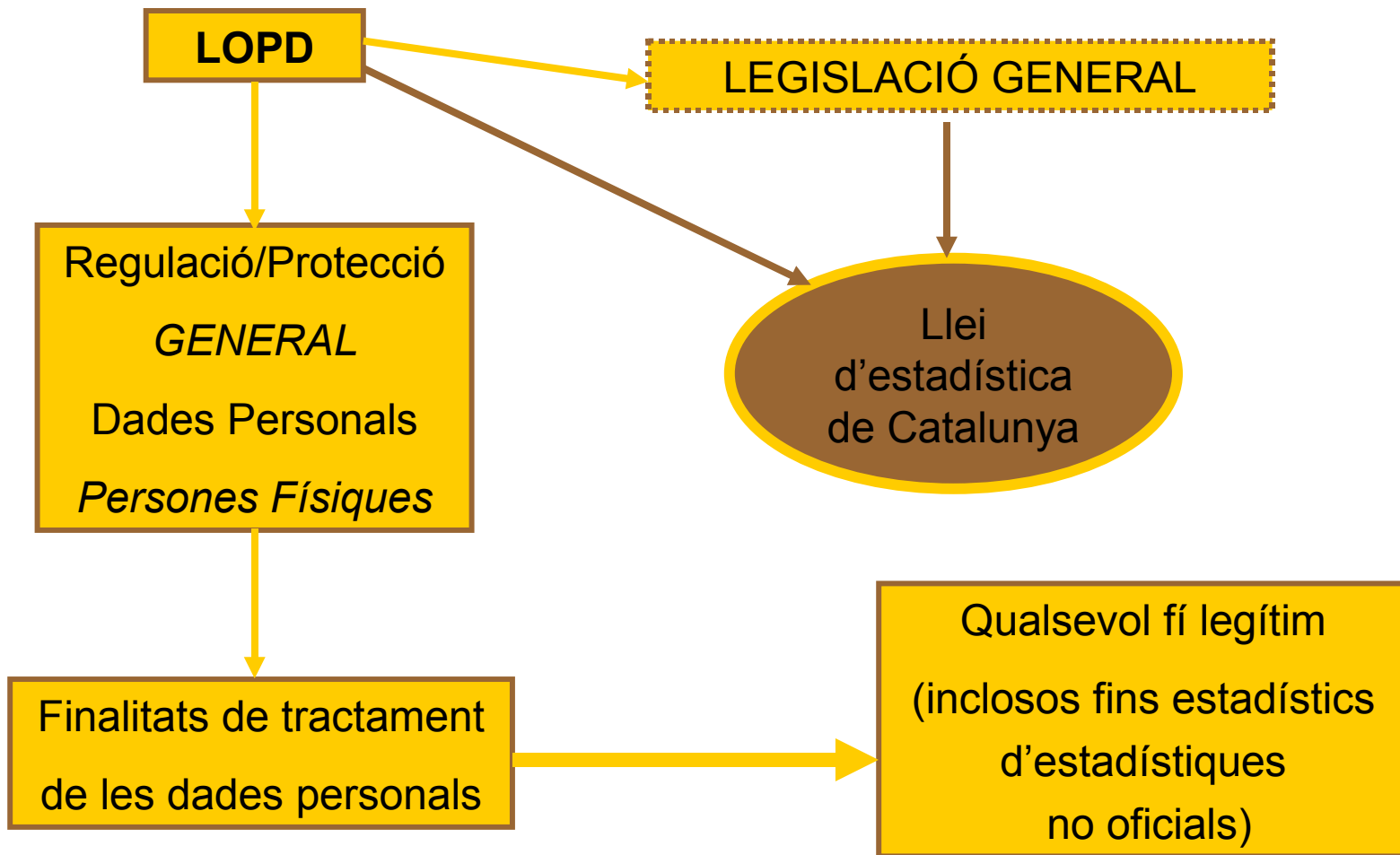
1.1 La interacció entre la LOPD i la LEC

Marc legal en protecció de dades personals

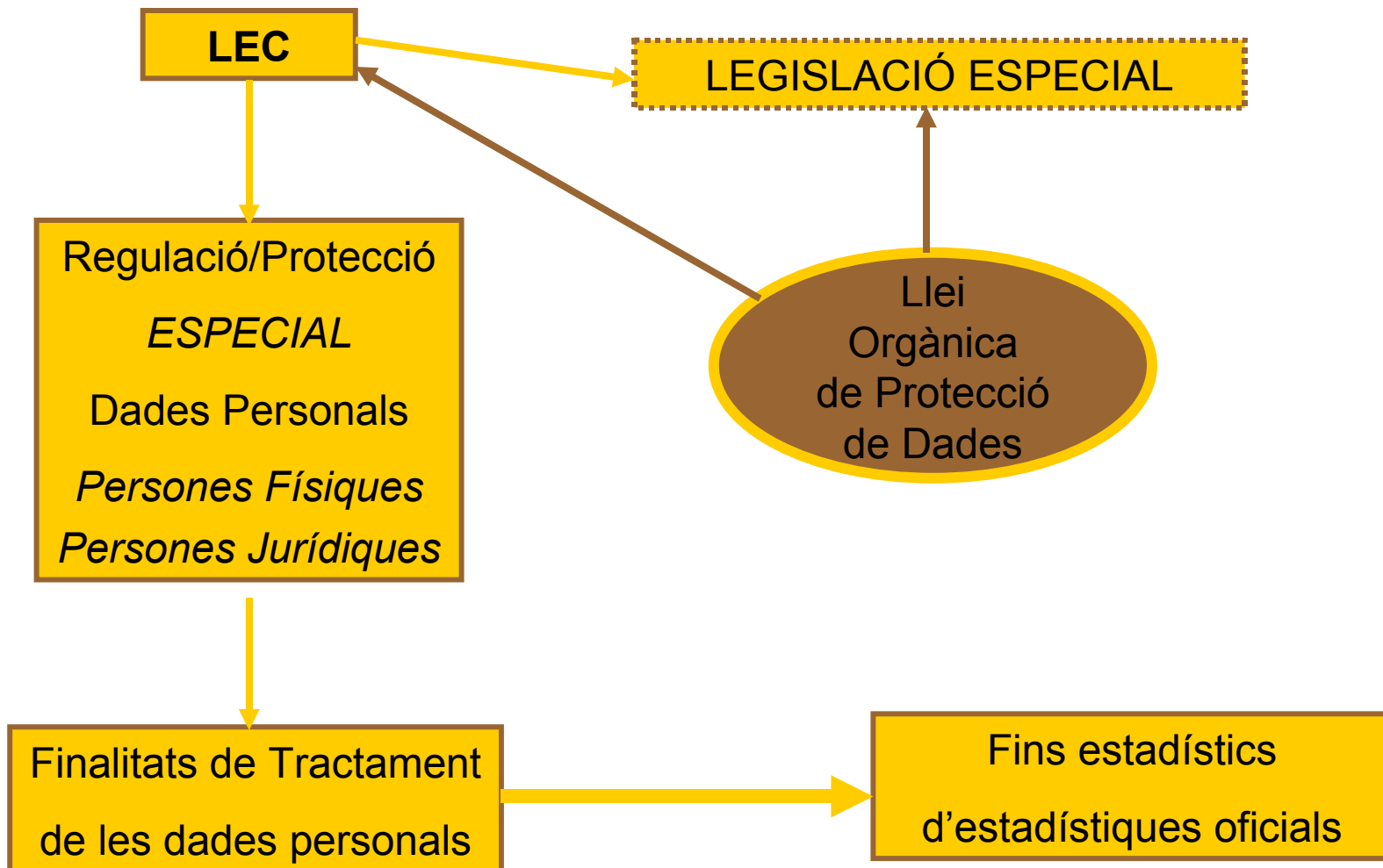
- **Llei Orgànica** 15/1999, de 13 de desembre, de Protecció de Dades de Caràcter Personal
- **Reglament** de mesures de seguretat dels fitxers automatitzats que continguin dades de caràcter personal (Reial Decret 994/1999, d'11 de juny)

Nou reglament en curs

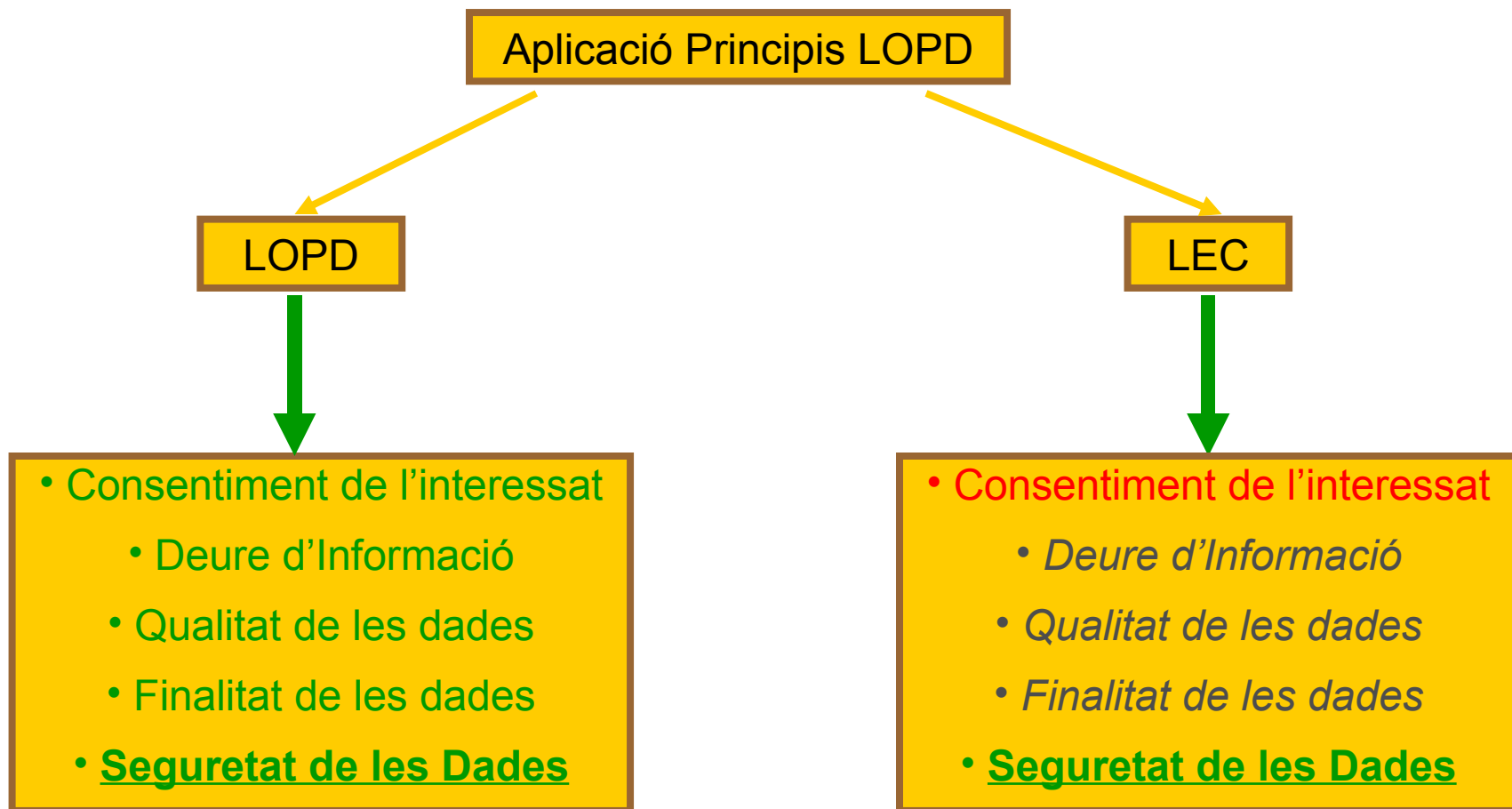
Aplicació de la LOPD i de la LEC (1/2)



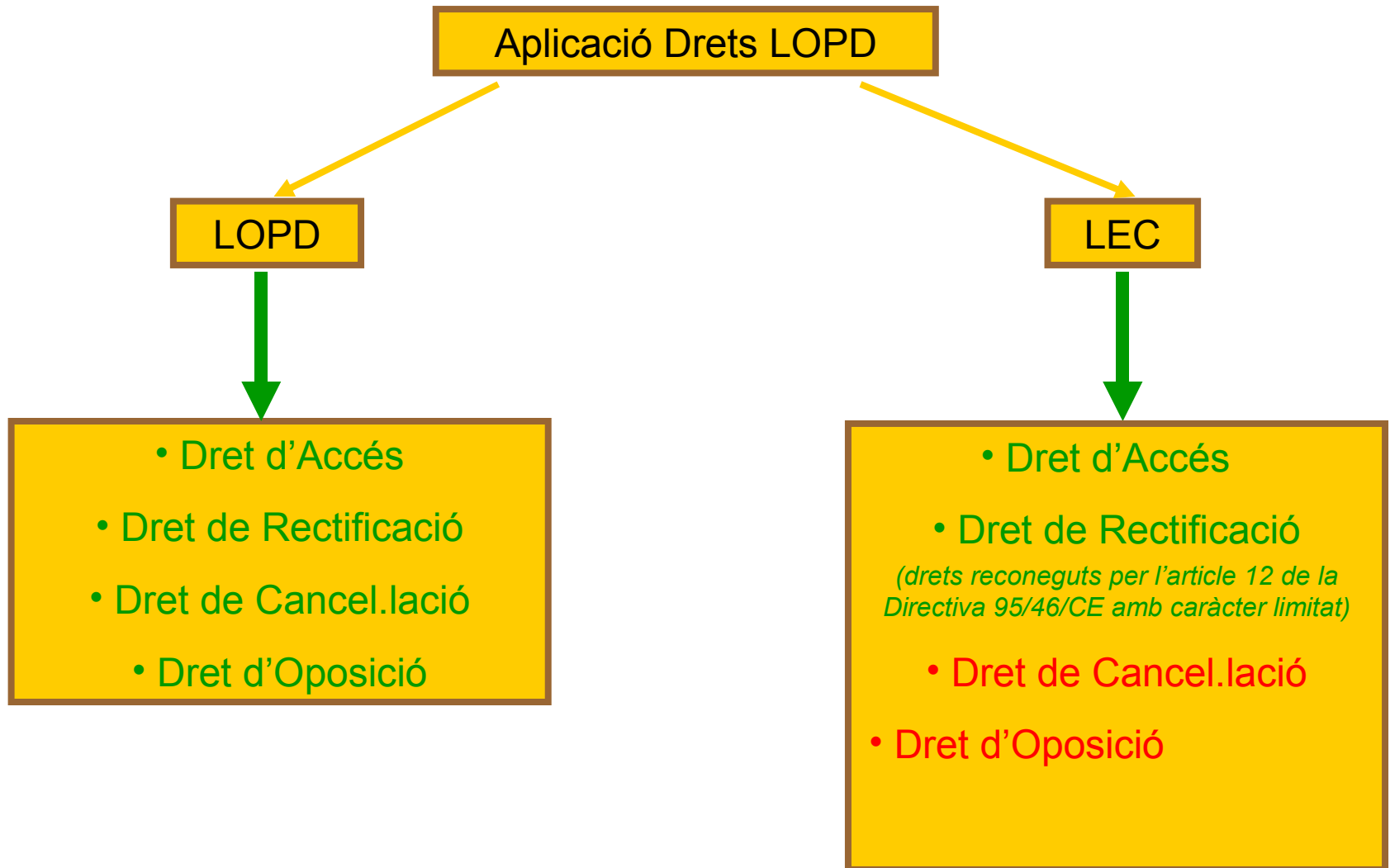
Aplicació de la LOPD i de la LEC (2/2)



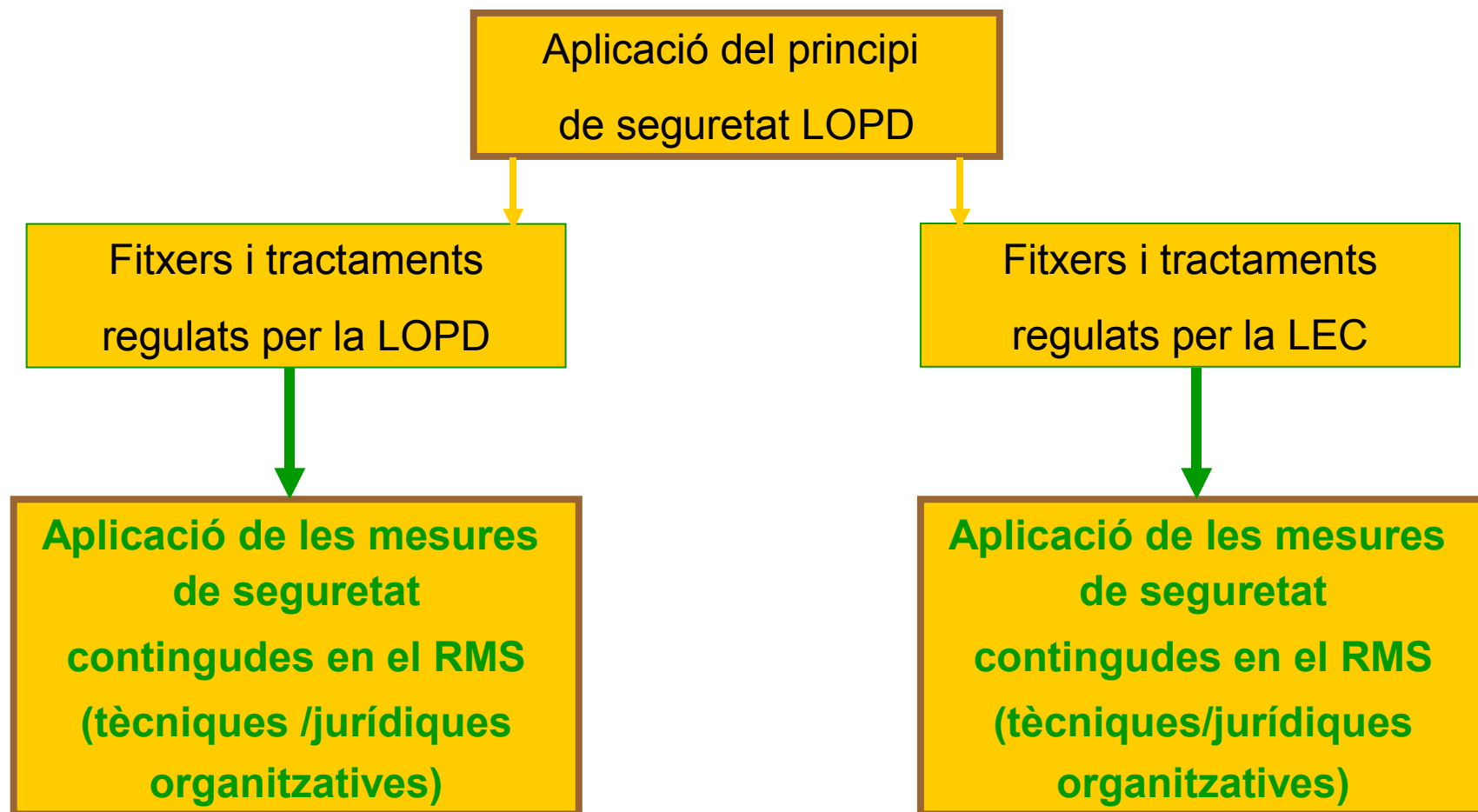
Diferències entre la LOPD i la LEC (1/3)



Diferències entre la LOPD i la LEC (2/3)



Diferències entre la LOPD i la LEC (3/3)



1.2 Marc legal i preceptes sobre secret estadístic (I)

□ Marc legal

- Llei d'estadística de Catalunya de 1998
- Llei del Pla estadístic de Catalunya 2006-2009
- Llei de la funció pública estadística de 1989
- Codi de bones pràctiques (UE): principi 5
- Reglament 831/2002 d'accés a dades per a recerca

□ Preceptes clau

- Doble compromís en la difusió i la confidencialitat
- Ús i accés a dades només amb finalitats estadístiques
- Cessió d'informació sotmesa a secret estadístic
 - entre organismes del SEC ≠ entre AA.PP.
 - investigadors / instituts de recerca



Marc legal i preceptes sobre secret estadístic (II)

- Excepcions i singularitats
 - Directoris amb informació de difusió general (*excepció*)
 - Renúncia al secret del titular de les dades (*excepció*)
 - Només dret d'accés i limitat als directoris estadístics
 - Criteri explícit en poblacions < 5.000 habitants
 - El cas del Registre de Població de Catalunya

- Algunes diferències amb altres sistemes estadístics
 - Estatal: detalls / terminis / "impossibilitat" de revelació
 - Codi europeu: divulgació dels mètodes SDC aplicats

2. Control de la revelació estadística: tecnologia

Mesures de seguretat i mètodes SDC

- Mesures de seguretat (Reglament de mesures de seguretat en el Reial Decret 994/1999 d'11 de juny): dispositius organitzatius i tècnics sobre gestió de suports, accessos, còpies i comunicacions
- Procediments pel control de la revelació estadística (SDC)
 - El risc de revelació estadística és una mesura de la probabilitat d'identificació de dades confidencials individuals per part d'un usuari, a través d'algun fitxer o taula que qualsevol organisme ha difós amb finalitats estadístiques (o d'altres)
 - No hi ha garantia absoluta de confidencialitat si es difon. El control garanteix riscos de revelació assumibles i pèrdues d'informació mínimes

Eines de preservació del secret estadístic: àmbits i mètodes SDC

Àmbits d'aplicació del control de la revelació estadística (SDC)

- Microdades o registres individuals
- Macrodades o dades tabulades

Famílies de mètodes SDC

- Supressió o reducció de dades (registres o cel·les d'una taula)
- Emmascarament, pertorbació o modificació de les dades

Eines de preservació del secret estadístic: àmbits (I)

- Exemple de fitxer de **microdades** o de registres individuals

	Identificadors directes		Variables identificadores*			Informació confidencial**				
	Nom	Adreça	I1	I2	I3	V1	V2	V3	...	Vm
Ind 1	Joan	c/....	I ₁₁	I ₂₁	I ₃₁	V ₁₁	V ₂₁	V ₃₁	...	V _{m1}
Ind 2	Josep	Av....	I ₁₂	I ₂₂	I ₃₂	V ₁₂	V ₂₂	V ₃₂	...	V _{m2}
Ind 3	Maria	c/....	I ₁₃	I ₂₃	I ₃₃	V ₁₃	V ₂₃	V ₃₃	...	V _{m3}
Ind 4	Anna	c/....	I ₁₄	I ₂₄	I ₃₄	V ₁₄	V ₂₄	V ₃₄	...	V _{m4}
...
Ind n-1	Olga	Pg...	I _{1n-1}	I _{2n-1}	I _{3n-1}	V _{1n-1}	V _{2n-1}	V _{3n-1}	...	V _{mn-1}
Ind n	Jordi	c/....	I _{1n}	I _{2n}	I _{3n}	V _{1n}	V _{2n}	V _{3n}	...	V _{mn}

* Edat, sexe, lloc naixement, professió / nombre treballadors, sector d'activitat, etc.

** Sou, malaltia, discapacitat / facturació neta, inversions, cost matèries primeres, etc.

Eines de preservació del secret estadístic: àmbits (II)

- Quan és necessari aplicar les tècniques de control de revelació estadística?
 - Microdades: sempre (són informacions individuals)
 - Eliminar identificadors directes
 - Eliminar/tractar variables identificadores
 - Eliminar/tractar variables sensibles (informació confidencial)

Eines de preservació del secret estadístic: àmbits (III)

- Exemple de dades tabulades/taula o **macrodades**

	X	Y	Z	Total
A	20	8	17	45
B	50	19	32	101
C	10	22	12	44
Total	80	49	61	190

Eines de preservació del secret estadístic: àmbits (i IV)

- Quan és necessari aplicar les tècniques de control de revelació estadística?
 - **Macrodades:** aplicació de les regles de sensibilitat
 - **Regla del valor llindar:** una cel·la serà sensible si la freqüència està per sota d'un determinat valor llindar. Taules de freqüències.
 - **Regla de dominància (n,K):** una cel·la serà sensible si les n contribucions més grans al valor de la cel·la constitueixen més del K% del valor total de la mateixa.
 - **Regla del percentatge (p):** una cel·la serà sensible si és possible aproximar alguna de las contribucions a la cel·la en més d'un p% del seu valor real.

$$\sum_{i=3}^c x_i \geq \frac{p}{100} x_1$$

← *El pitjor cas: la segona major contribució intenta estimar la primera.*

Eines de preservació del secret estadístic: mètodes SDC (I)

		Tipus de dades	
		Microdades	Macrodades
Tipus de tècnica	Supressió	Anonimització Mostreig Restricció segons N Recodificació Top/Bottom coding Supressió	Supressió Reducció taula/recodificació
	Enmascarament	Pertorbació aleatòria Data Swapping Pertorbació aleatòria amb compensació Microagregació	Arrodoniment Arrodoniment amb compensació Pertorbació aleatòria Pertorbació aleatòria amb compensació

Eines de preservació del secret estadístic: mètodes SDC (V)

- Tècniques de supressió o reducció de (macro)dades
 - Supressió de cel·les: consisteix en no proporcionar la informació d'aquelles cel·les que siguin especialment conflictives (sensibles), ja sigui perquè la freqüència que representen és menor a un nombre determinat (regla de l'umbral) o perquè uns pocs individus representen un percentatge elevat del total de la cel·la (regla de la dominància). La supressió d'una cel·la (supressió primària) suposa la supressió d'altres cel·les de la mateixa taula (**supressió secundària**) per tal d'evitar la deducció del valor de la cel·la primerament suprimida.
 - Reducció de la taula i recodificació: consisteix en agregar les categories d'una variable a fi de no donar informació tan detallada, és a dir, reducció de la dimensionalitat de la taula.

Eines de preservació del secret estadístic: mètodes SDC (i VI)

□ Supressió de cel·les secundàries

Taula Original

	ACT. I	ACT. II	ACT. III	Total
A	20	8	17	45
B	50	19	32	101
C	10	22	12	44
Total	80	49	61	190

Críteris de sensibilitat



Supressions primàries

	ACT. I	ACT. II	ACT. III	Total
A	20	8	17	45
B	50	19	32	101
C	10	22	12	44
Total	80	49	61	190



Supressions secundàries

	ACT. I	ACT. II	ACT. III	Total
A	20	8	17	45
B	50	19	32	101
C	10	22	12	44
Total	80	49	61	190



Taula publicada

	ACT. I	ACT. II	ACT. III	Total
A	20	8	17	45
B	.	19	.	101
C	.	22	.	44
Total	80	49	61	190

3. Criteris sobre preceptes i mètodes: opcions i reptes

- a. Adopció de mètodes i criteris
 - Opcions **metodològiques**. Implantació
 - Adopció i socialització de **criteris** (interna i externa)

- b. Adequacions de noves tecnologies de difusió
 - Accés a microdades **originals** \ Cubs d'informació (i.e. *Beyond 20/20*)

- c. Adequacions legals/institucionals
 - Regulació del **Registre de Fitxers Estadístics**
 - ... **Reglament** de la llei d'estadística (o nova llei?) ...

3.1. Implantació de mètodes i socialització de criteris en SDC

- Adopció de mètodes SDC
 - modificació vs. emmascarament = f (anàlisi)
 - submostres controlades vs microdades originals
 - disseny de bases de dades interactives
 - taules jeràrquiques / taules vinculades

- Referents sobre opcions metodològiques
 - Àmbit de microdades
 - Entorn de dades tabulades

3.1. Implantació de mètodes i socialització de criteris en SDC

- Adopció de criteris
 - Dades censals o per mostreig: aplicació de SDC?
 - Microdades: difusió mínima per incapacitat o compromís
 - Titularitat/propietat de les dades: regulació en convenis
 - Investigador: com encarregat de tractament / usuari final
 - **Fitxers compartits amb INE/Ad. central: criteris diferents?**

- Socialització de criteris
 - Comitè SDC
 - Economies d'escala del Document de Seguretat de fitxers LOPD
 - Efectes sobre projectes tècnics, convenis, equipaments, ...
 - CATIC: CCE i/o investigadors, productors i experts

3.1. Accés segur a dades originals i altres modes de difusió

□ Modalitats

- Accés segur “on site” (referent: CBS)
- Accés segur remot (referent: StatCan)
- Difusió de cubs d’informació (i.e., *Beyond 20/20*)

□ Elements a considerar

- Adequació de la legalitat (art. 30 LEC ... i altres?)
- Organització interna, complexitat tècnica i costos materials/humans

3.2. Criteris sobre accés a microdades: el cas Idescat (I)

- Tipologia de sol·licitants d'accés a microdades
- Nivells d'accessibilitat a la informació confidencial
- Tractaments i protocols d'aplicació als sol·licitants



Criteris corporatius sobre accés a microdades (II)

□ Tipologia de sol·licitants

- Investigadors tipus I: investigadors encarregats del tractament de dades
- Investigadors tipus II: investigadors que actuen només com usuaris finals
- Organismes corresponsables/col·laboradors d'activitat estadística oficial
- Altres organismes de l'administració pública (estatal, autonòmica o local)
- Usuaris genèrics: sol·licitants no inclosos en les categories anteriors

□ Nivell d'accessibilitat

- Microdades originals: investigadors tipus I i organismes corresponsables
- Microdades amb risc d'identificació indirecta: investigadors tipus II
- Microdades sense risc d'identificació indirecta: resta de sol·licitants

Criteris corporatius sobre accés a microdades (i III)

- **Exigències als sol·licitants** (paràmetres rellevants)
 - Adequació de la actuació/recerca a l'accés imprescindible a microdades
 - Pla de tabulació executat per l'Idescat o altres (co)responsables
 - Disponibilitat d'altres fonts estadístiques i administratives d'accés públic
 - Protocols segons nivells d'accessibilitat: de més a menys exigències
 - Compromisos d'emmagatzament, tractament i redifusió:
 - Control del pla de tabulació /resultats que difondria el sol·licitant
 - Advertiments sobre la qualitat de les dades, drets d'autor i similars
 - Conformitat a eventuais inspeccions o controls per part de l'Idescat

3.3. El futur Registre de Fitxers Estadístics del SEC

Disposició final primera (Llei 2/2006 del PEC 2006-2009)

Creació del Registre de Fitxers Estadístics

1. Es crea el Registre de Fitxers Estadístics i s'adscriu a l'Institut d'Estadística de Catalunya.
2. El Registre de Fitxers Estadístics ha d'inventariar els fitxers amb finalitats exclusivament estadístiques, referits tant a persones físiques com jurídiques, que les institucions i òrgans del Sistema Estadístic de Catalunya creïn exclusivament per a elaborar estadístiques d'interès de la Generalitat de Catalunya.
3. Per decret del Govern, a proposta del conseller o consellera d'Economia i Finances, s'ha de regular el Registre de Fitxers Estadístics i s'ha d'establir el procediment d'inscripció dels fitxers corresponents i la informació que s'hi ha d'incorporar.