A shared portal for the research output of the universities of Catalonia

Lluís Anglada & Sandra Reoyo
CBUC (CSUC)

14th SELL Meeting
Firenze, May 23rd

Outline

1. Background, objectives and means
2. Work packages
   1. Elements
   2. Identifiers
   3. Data flow
   4. Portal building
3. Work to be done & challenges

La situació el 2012

- Current situation
  - CBUC promotes IR since 1999
  - in general, in CBUC universities, libraries plays an active role promoting OA and the quality of CRIS data
  - Some universities (UPC i UPF) already have research portals
- Opportunities
  - The new consortia (CSUC) is more influent than CBUC was
  - Research data are becoming central in research
  - There are new standards and protocols that help interoperability between IR and CRIS

What, why & how

What

- A portal (= unique place) where to find the research outputs of the Catalan research system

Why

- To increase the visibility of the research done in Catalonia
- To foster OA
- To increase interoperability between data

How

- Taking advantage of the leverage work previously done
  - In IR, CRIS and statistical data (Uneix)
- The central idea: the works done for the portal will improve local IR and CRIS
  - Standards and protocols will help the interoperability of the local data
- Following international best practices
  - Narcis / Holland; HKU Scholars Hub / Hong Kong;
- Big effort in communication
  - 28 meetings in 6 months

Work packages

- Selection and definition of the elements that will appear in the portal
- Agree the identifiers that will be essential to avoid duplications (especially between researchers)
- Data flow: how the elements will be exported (form CRIS) and imported (to the portal)
- Portal building

Outline

1. Background, objectives and means
2. Work packages
   1. Elements
   2. Identifiers
   3. Data flow
   4. Portal building
3. Work to be done & challenges

2.1 Apartats

Universitats

- Departaments i Instituts
- Grups de recerca
- Investigadors (PDI + PI)
  - Projectes de recerca
  - Publicacions (Articles + Llibres + Tesis)

Elements

Universitats	Departaments i Instituts	Grups de recerca	Projectes de recerca	Investigadors	Publicacions
Nom	Nom	Nom	Títol	Nom	DOI
Sigles	Sigles	Sigles	Codi	Cognom/s	Handle
Adreça	Adreça	URL	Programa	ORCID	Títol
URL	URL	a/e	Data d'inici	a/e	Autor/a (amb ORCID)
a/e	a/e	Pertany a	Data de fi del projecte	Universitat	Data de publicació
Situació (Google maps)	Pertany a	Codi SGR	Nom i Cognom/s dels investigadors (amb ORCID)	Departament/Institut	Publicat a
Telèfon	Situació (Google maps)	Data de creació			Publicat per
Fax	Telèfon	Àmbit de recerca			Tipus de document
		Nom i Cognom/s investigador principal (amb ORCID)			
		Nom i Cognom/s dels investigadors membres (amb ORCID)			

2.2 Identifiers (ORCID)

1. Selection of identifiers
   - Decision based in a CBUC report: Sistemes d'identificació unívoca d'investigadors / Àngel Borrego
   - Debated in a working group; approved in a meeting of Vice Provosts 24.07.13
2. Technical work
   - Modify all the local CRIS in order to allow to load the ORCID identifier
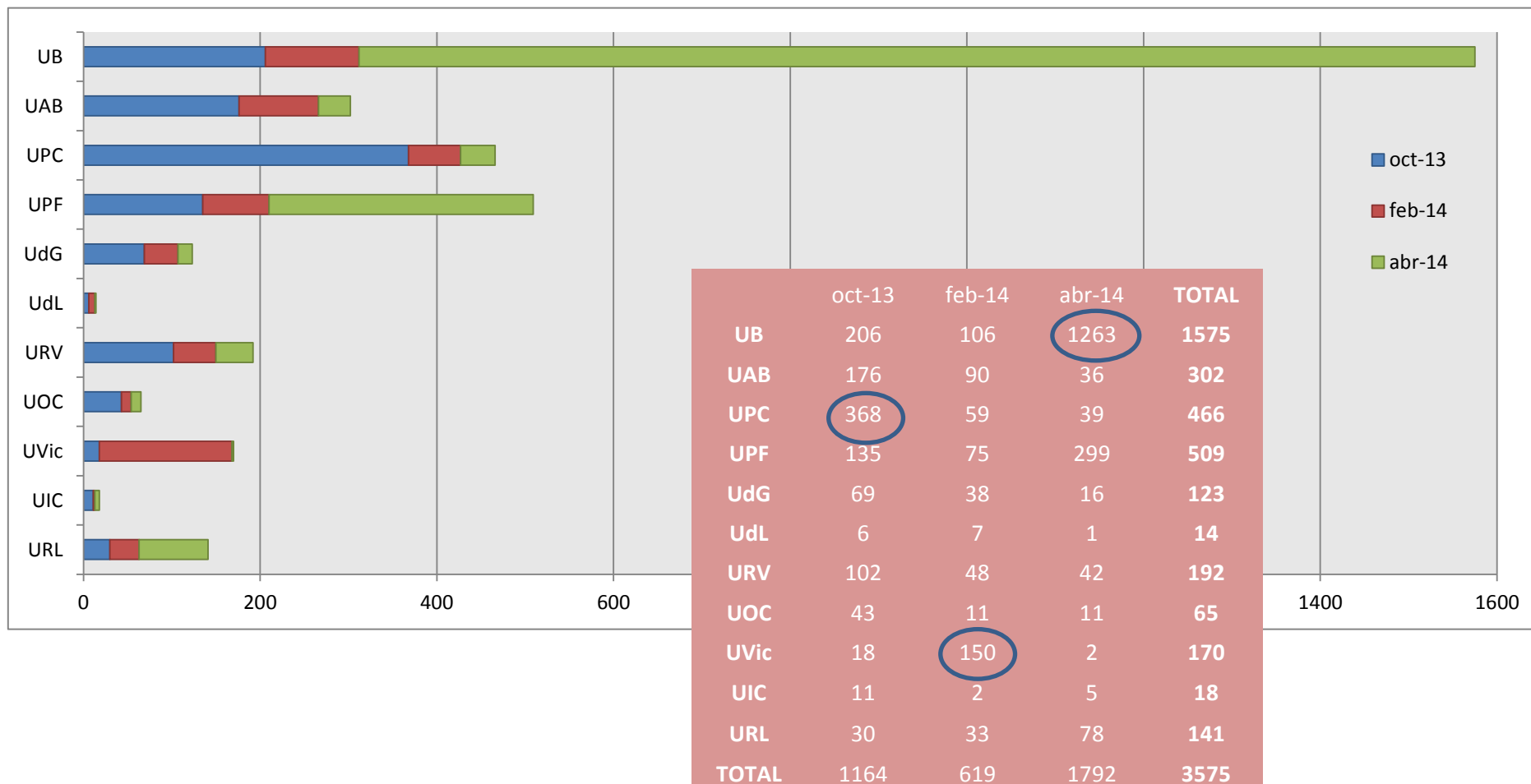   - Promotion of ORCID id in other working groups: repositories, CCUC, Mendeley…
3. ORCID diffusion
   - We studied the ORCID apps, to create ORCID id automatically, but we decided not to use it
   - Merchandising, translations, videos...
   - Vice provost approved a 'good practices' document in order to promote the creation and usage of ORCD ids
4. Political work
   - UB (the biggest university) mandate for an ORCID id in some process related with research assessment
   - We are trying to do the same at Catalan government level
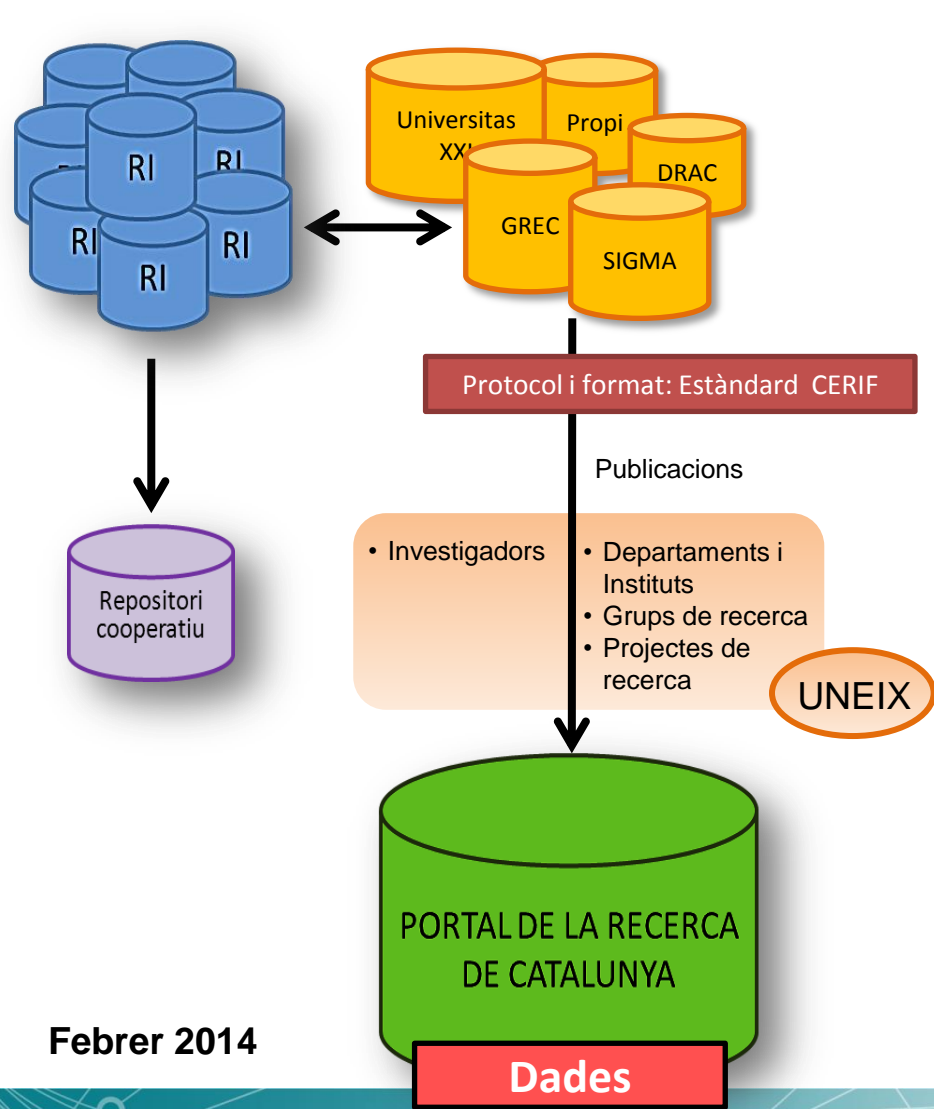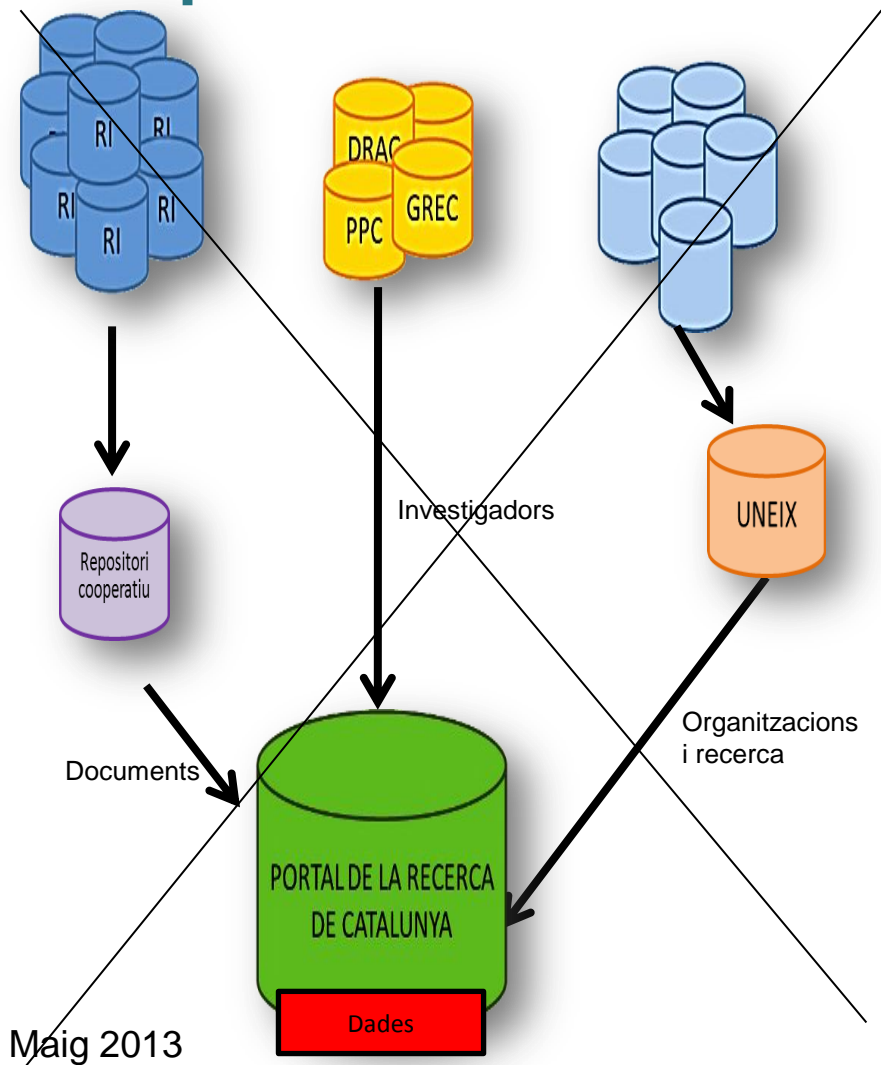
Evolution of Catalan researchers with ORCID*

	oct-13	feb-14	abr-14	TOTAL
UB	206	106	1263	1575
UAB	176	90	36	302
UPC	368	59	39	466
UPF	135	75	299	509
UdG	69	38	16	123
UdL	6	7	1	14
URV	102	48	42	192
UOC	43	11	11	65
UVic	18	150	2	170
UIC	11	2	5	18
URL	30	33	78	141
TOTAL	1164	619	1792	3575

* Dades proporcionades per ORCID - Investigadors donats d'alta amb correu electrònic de la universitat

CSUC

2.3 Data flow, protocols, sources and formats

1. Where will data come from?
   - CRIS from each university as a unique source
   - CRIS will be upated from: IR, staff database, external providers databases, etc.
2. We need to sign a memorandum of understanding for personal data protection
   - We need lawyers!!!
3. What protocols and schemas are we going to use?
   1st: just sample of 20 researches in XLS format
   2nd: all data in XLS format
   3rd: CERIF-XML file (GrandIR is writing the specification based on OpenAire-CERIF guidelines).
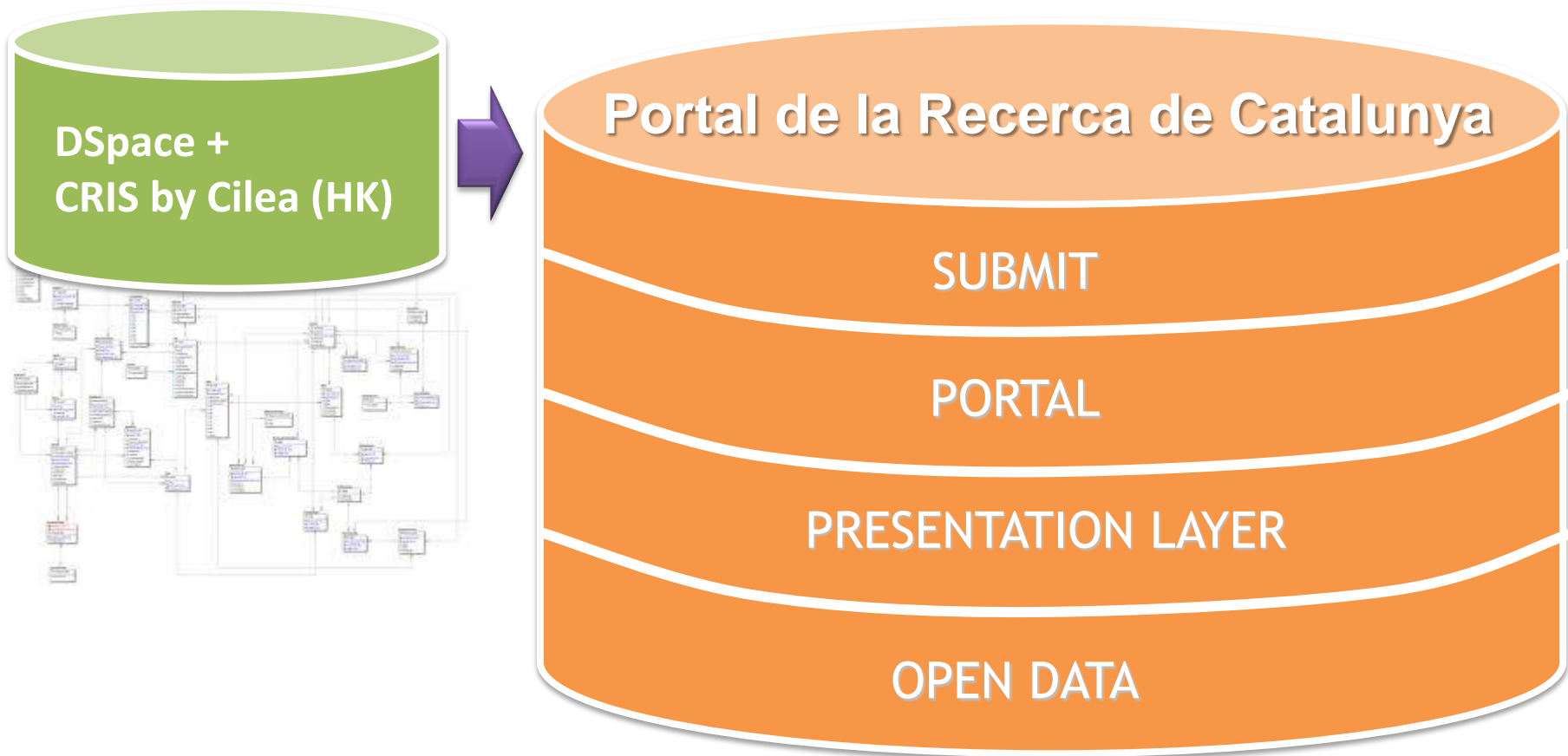   4th: full CERIF-XML through OAI-PMH protocol

Data flow, protocols, sources and formats

2.4 Portal Building

- Based on DSpace-CRIS of CILEA (like Hong Kong University)

- Main challenges (to adapt/develop)
  - From one institution to multi-institution
  - From submit contents to harvest from local CRIS instances
  - Massive import mechanisms are needed (XML-CERIF….)

Portal building

DSpace +
CRIS by Cilea (HK)

Portal de la Recerca de Catalunya

SUBMIT
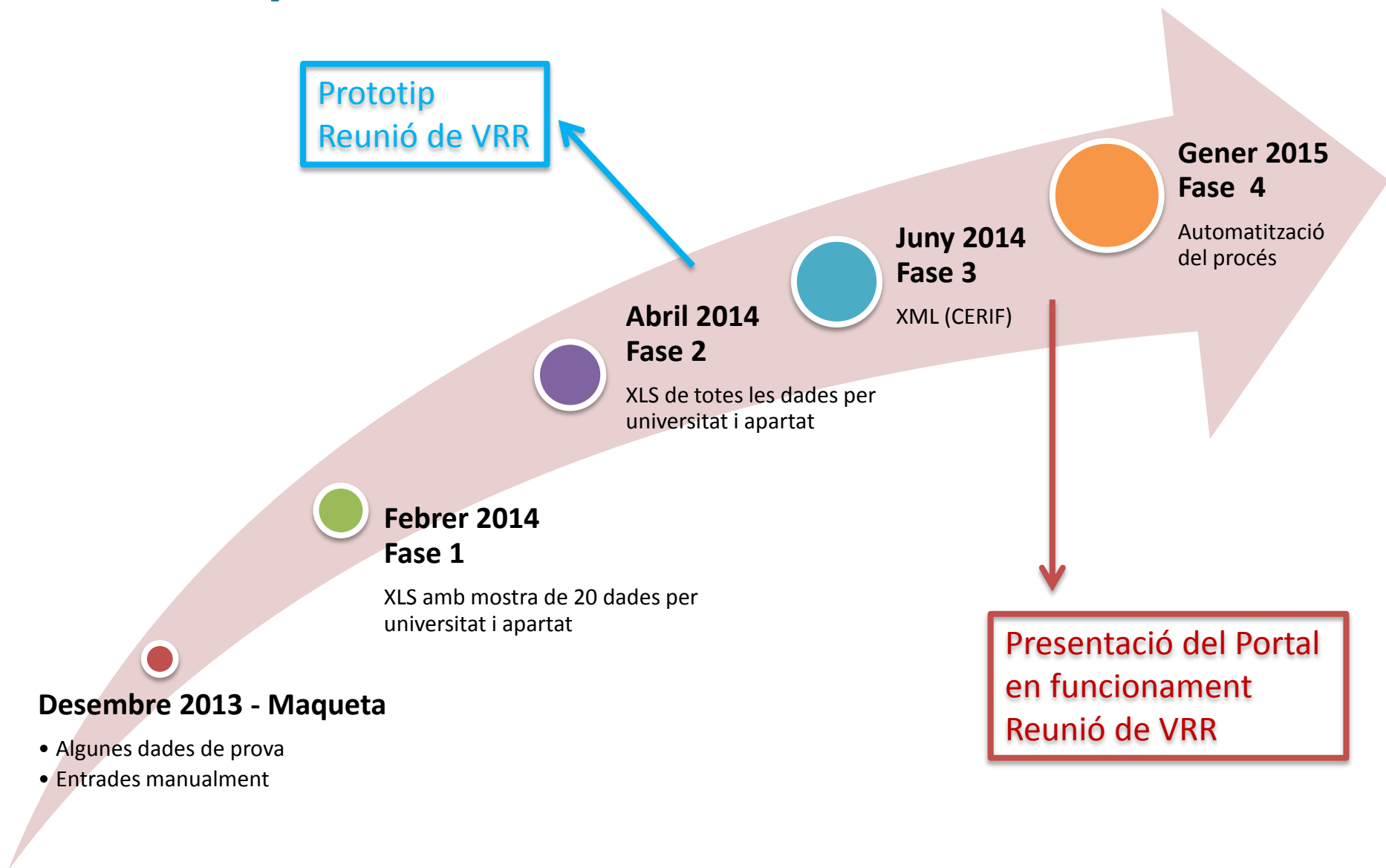
PORTAL

PRESENTATION LAYER

OPEN DATA

Outline

1. Background, objectives and means
2. Work packages
   1. Elements
   2. Identifiers
   3. Data flow
   4. Portal building
3. Work to be done & challenges

CSUC

Main achievements

- We have a unique objective and a good working team
  - People from ≠ universities and ≠ services
- Agreement: to use ORCID for researchers
- Already done
  – We succeed to export 20 complete data records from 11 universities (using 5 different CRIS)
  – All the CRIS systems already have a field for ORCID
  – A good programme selected
  - Adopted by EUROCRIS as repository because CERIF compliance

Proposta de fases i calendari

Prototip
Reunió de VRR

**Gener 2015
Fase 4**

Automatització
del procés

**Juny 2014
Fase 3**

XML (CERIF)

**Abril 2014
Fase 2**

XLS de totes les dades per
universitat i apartat

**Febrer 2014
Fase 1**

XLS amb mostra de 20 dades per
universitat i apartat

Desembre 2013 - Maqueta

• Algunes dades de prova
• Entrades manualment

Presentació del Portal
en funcionament
Reunió de VRR

CSUC

Work to be done & challenges

- More meetings
  - Working group and subgroups
- ORCID ids implementation
- MoU for personal data
- Data exportation
  - Excel
  - XML-CERIF
- Hard work to built the portal
  - Finish the prototip with data sample
  - Ingest the full data of all institutions
  - Design and build the user interfaces
  - Develop the CERIF-XML import mechanisms
  - Thing about depuration data mechanisms