# The Open Science Experience of the Rare Disease Community

Segolene Ayme

[Segolene.ayme@icm-institute.org](mailto:Segolene.ayme@icm-institute.org)

Institut du Cerveau– Paris, France

With the contribution of Marc Hanauer and Charlotte Rodwell from Orphanet

Neuro Open Science Workshop – 15 January 2019

# Sharing Data and Knowledge
# is a must in the field of Rare Diseases

- Direct difficulties
  - Knowledge is scarce
  - Experts are rare

- Derived obstacles
  - No single centre gather enough data to conduct clinical research
  - Funding is difficult to obtain

- Targets for action
  - Increasing awareness is a necessity
  - Every piece of data is precious
  - Stakeholders have interests in common
  - Society should provide solutions based on the equity principle

ICM

# Implemented Solutions so far

- Repository of knowledge: Orphanet -1997

- Incentives for Industry : the Orphan Drug Regulation - 2000

- Funding of rare disease registries (national and EC) - 2004

- Funding of multidisciplinary networks for research (national and EC) - 2004

- International Rare Disease Research Consortium-IRDiRC - 2012

- National Plans for Rare Disease in European countries - 2015

- Establishment of Centres of Expertise at country level

- European Reference networks of Centres of Expertise- ERN-2018

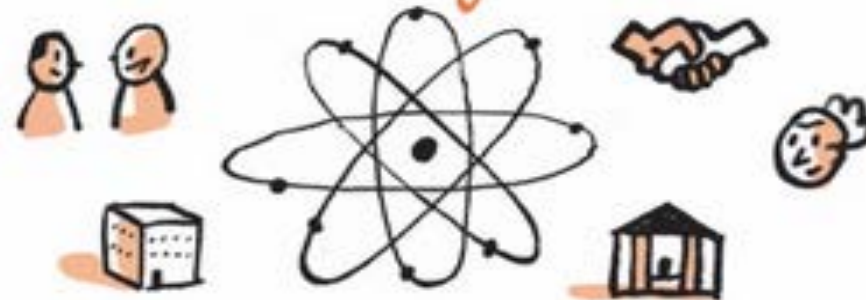# The ecosystem matters
# It is in place for Rare Diseases

# The Experience of Orphanet
## www.orpha.net
## www.orphadata.org

# Orphanet is a Curated Relational Database
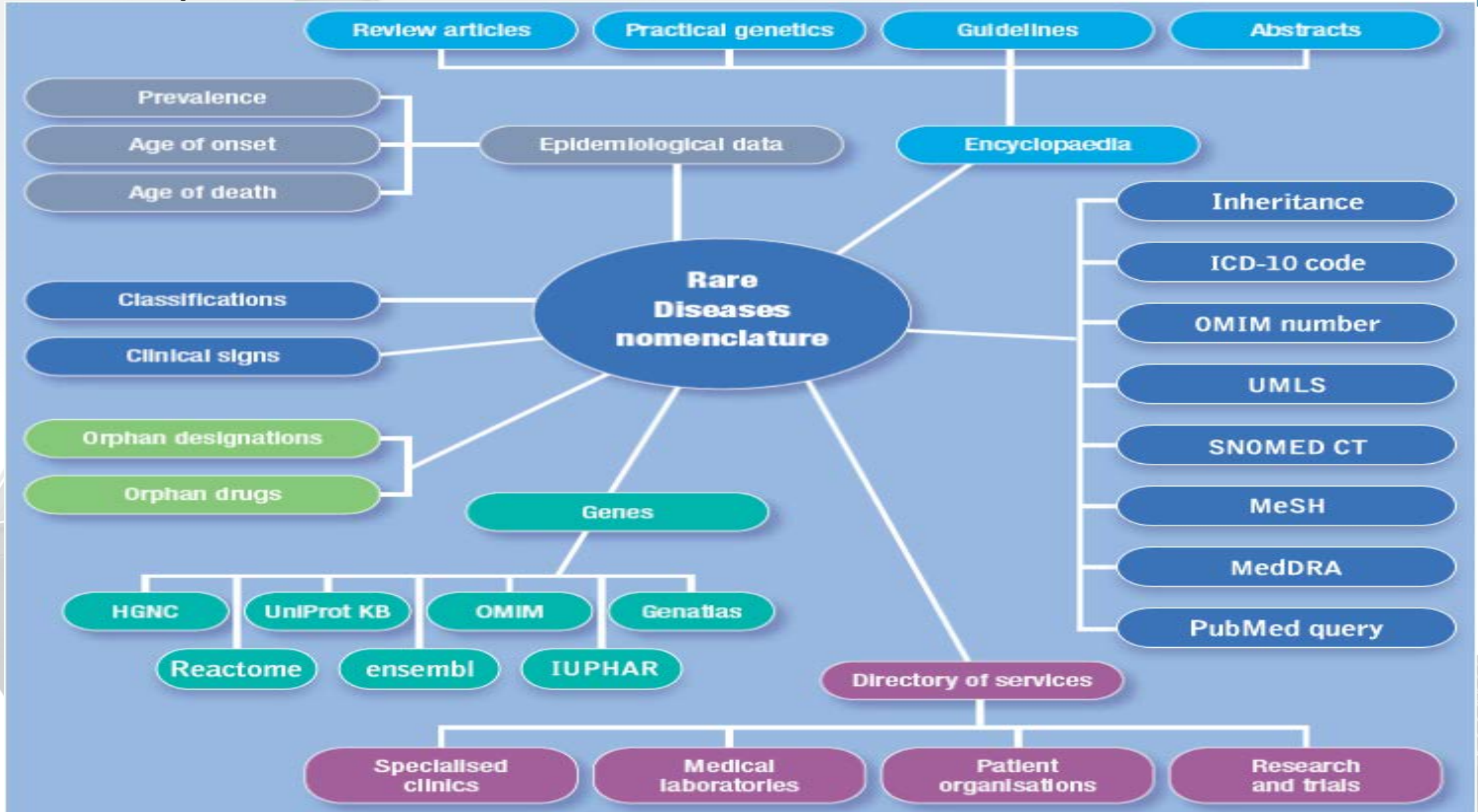
# From Orphanet to Orphadata in June 2011

- Orphanet provides access to knowledge and data sources on 7,000 single rare disease entities to 1 Million monthly users

- Questions:
  - How to boost Research & Development in the field ?
  - how to unify the community and provide tools for interoperability ?

- Response: Orphadata.org
  - Make the whole Orphanet dataset accessible, either freely, or on request

# ORPHADATA:an Elixir Core Resource

- License
  - [Commons Attribution-NoDerivatives 4.0 International Licence](#) 2011-2018
  - free to copy, distribute, display and make commercial use of the data in all legislations, providing credit attribution to Orphanet- No distribution of a modified version, unless authorized by Orphanet
  - Creative commons by Orphanet since January 2019
- Free access products
  - Inventory, classification, linked genes, ontology
- On Request
  - All other tables: signs and symptoms, epidemiology, drugs in development or on the market, registries, research projects, diagnostic labs, centres of expertise, abstracts, encyclopedia
  - Free of Charge for Academics
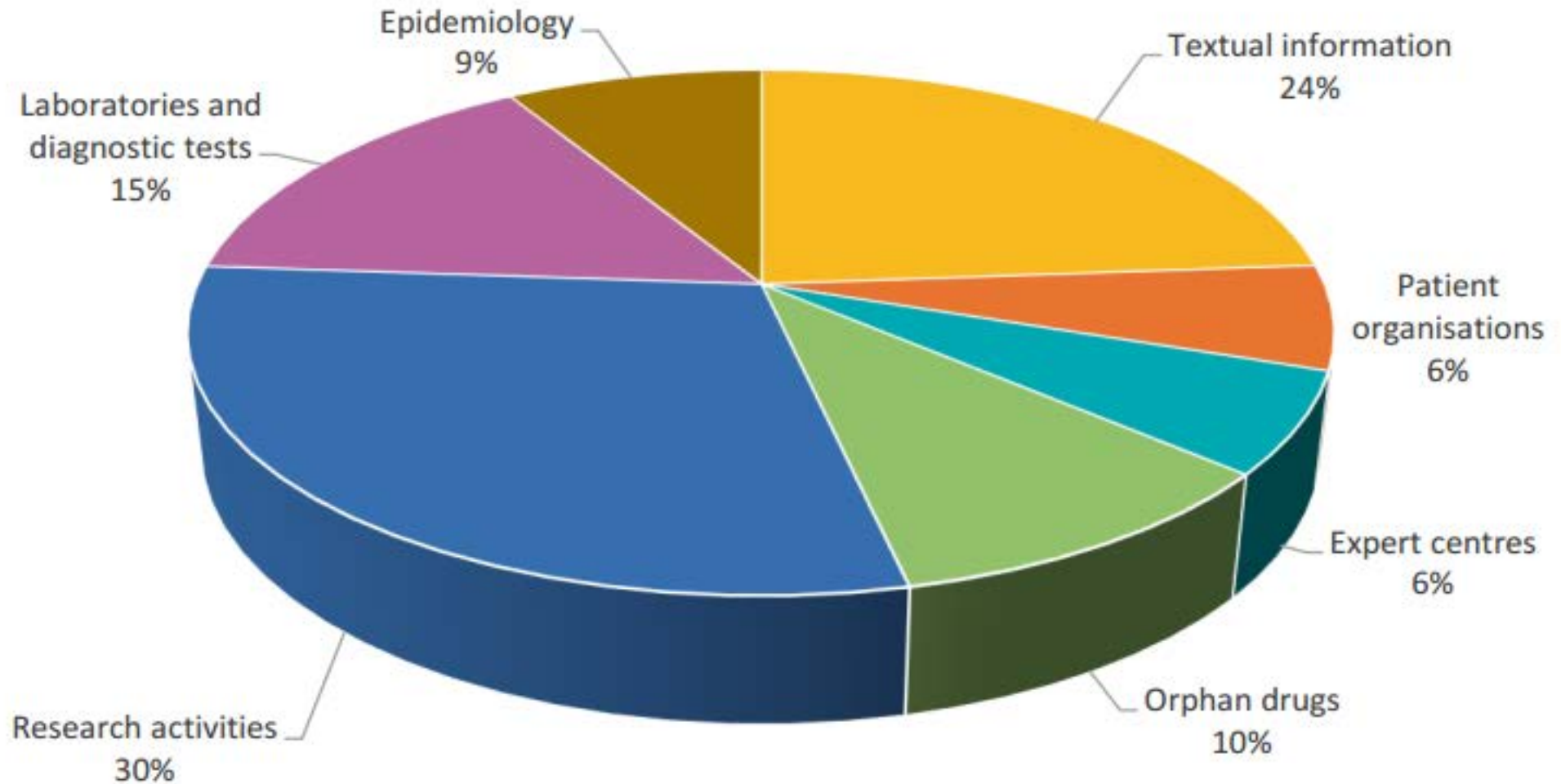  - For Fees for Industry: 120 K Euros/year

# Number of Dowloads of Orphadata products 2011- 2017
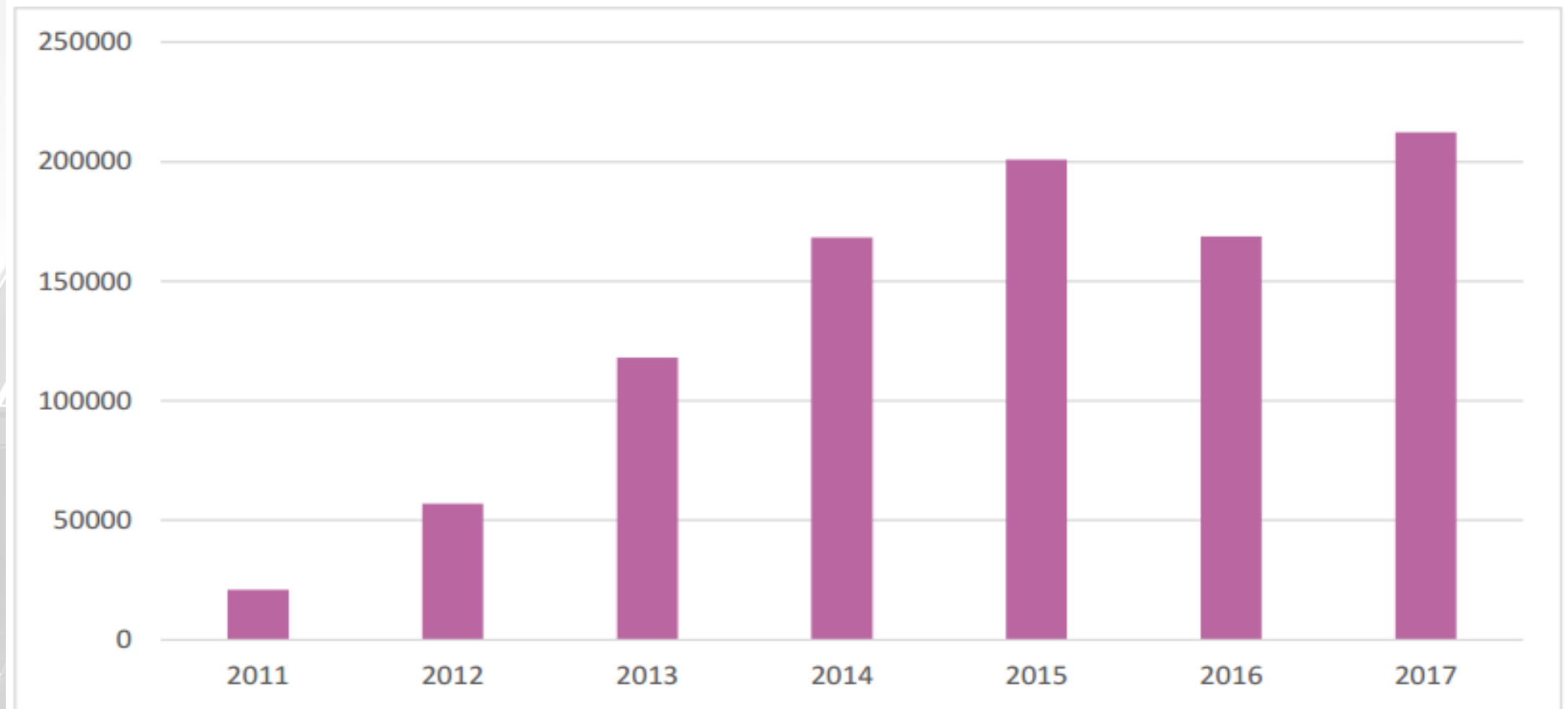
# Distribution of 209,028 Downloaded Free Access Products in 2017 (JSON format)
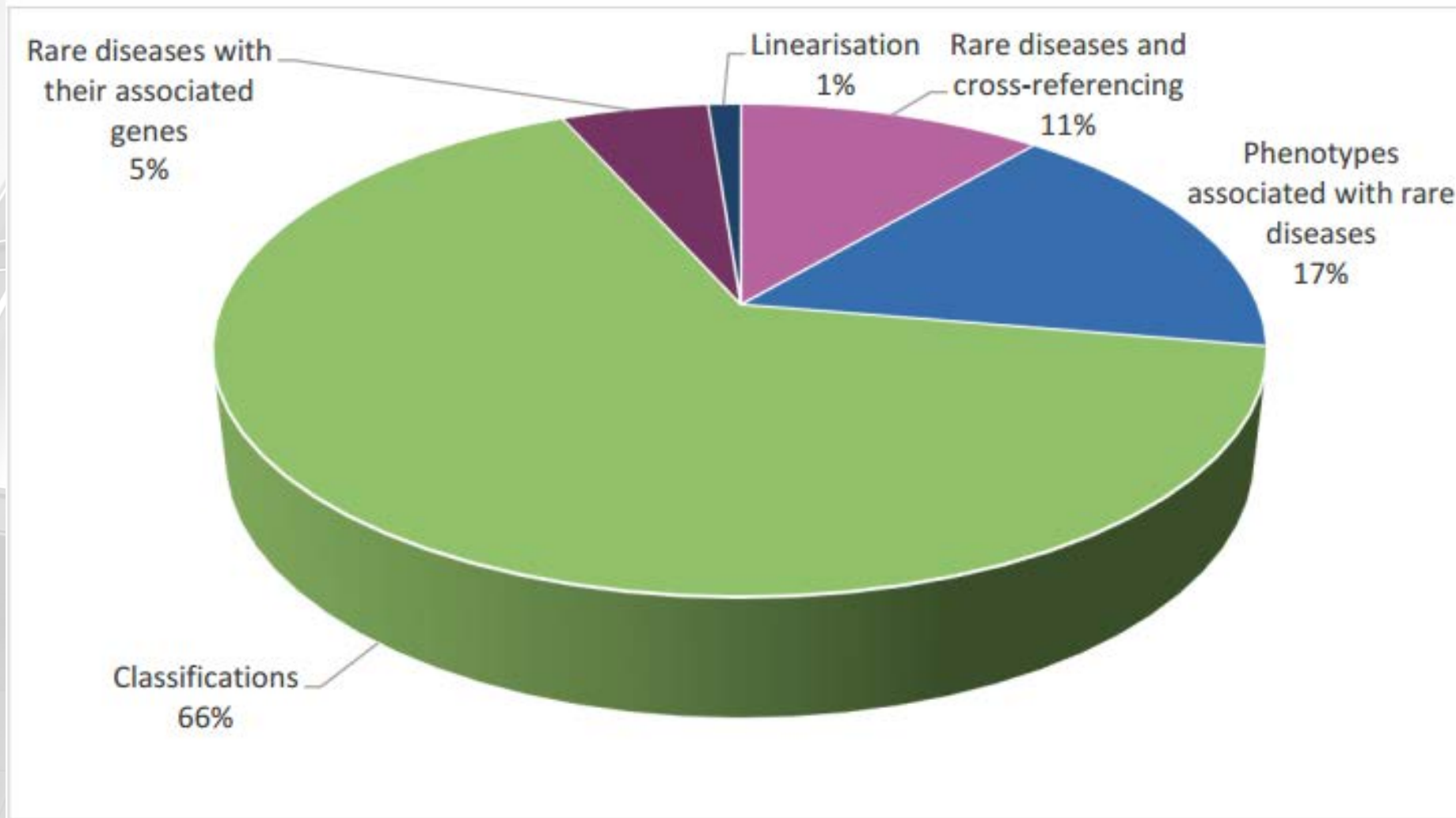
# Distribution of 3,243 Downloaded On Demand Products in 2017

# Impact of Orphadata on Publications

## Resource name mentioned in Europe PMC (citation of resource name)

- **2014      75 articles**
- **2015     111 articles**
- **2016      93 articles**
- **2017     142 articles**
- **2018     111 articles**

# Lessons from ORPHADATA

- Massive use of the dataset
- Establishment of the Orphanet nomenclature as the International standard (adopted by WHO for ICD 11)
- Establishment as and International infrastructure: IRDiRC and Elixir
- Many more collaborations + invitation to join EC funded and NIH funded grant applications
- Too complicated to respond to all academic requests
  - more should be directly accessible without DTA
  - Necessity to trace the users for activity reports
- Prices are too high and process (through INSERM-Transfert) too complicated
- Best approach: and API management system to charge above a given threshold –to come

ICM

# Business models proposed so far

Check for updates

RESEARCH ARTICLE

REVISED Funding knowledgebases: Towards a sustainable funding model for the UniProt use case [version 2; referees: 3 approved]

Chiara Gabella ⓘ, Christine Durinx ⓘ, Ron Appel

ELIXIR-Switzerland, SIB Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland
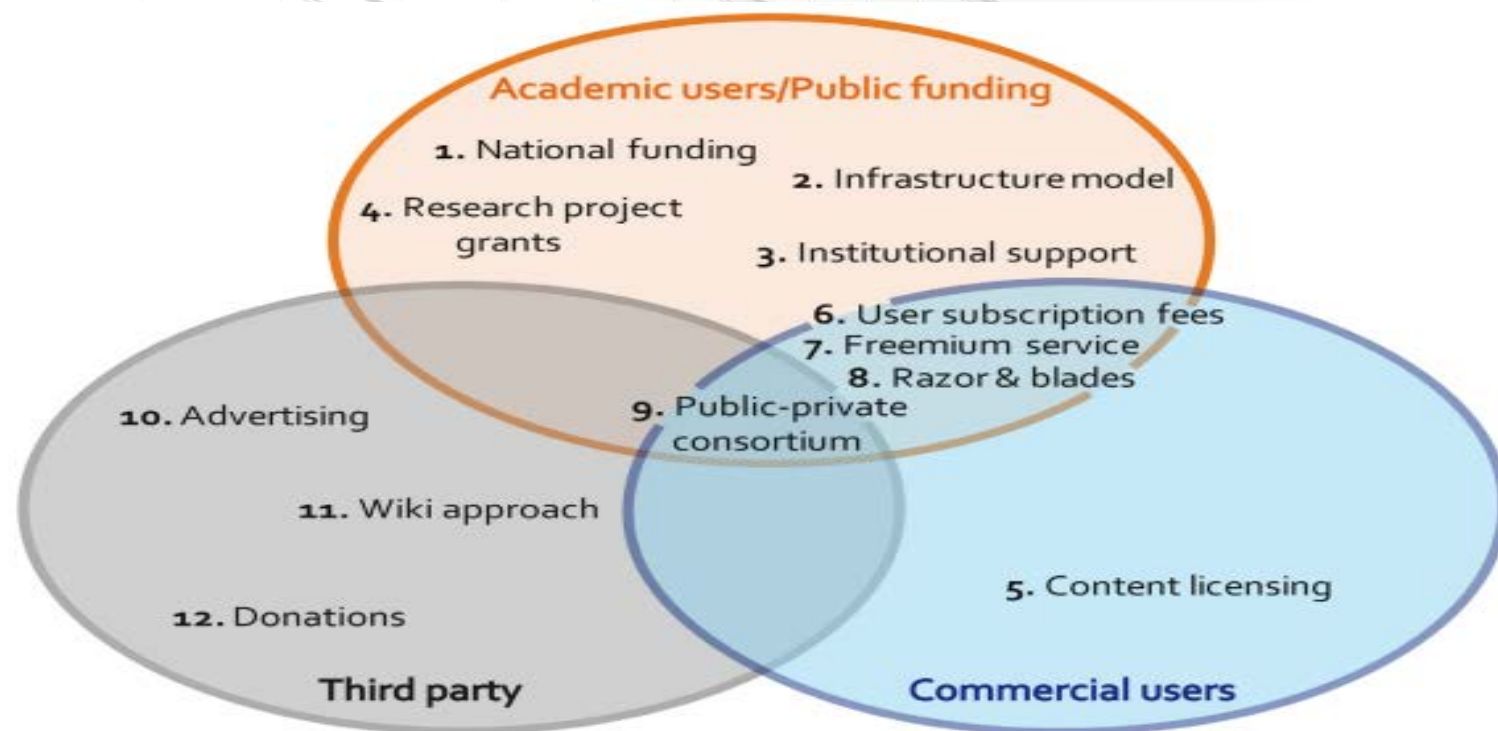


**Figure 1. Funding models sources.** The 12 considered models are represented depending on the origin of the revenues.

# Funding models for public databases proposed by UniProt

| # | Name of the model | Compatible with open access? | Potential for equity of users or institutions | Stability forecasted over time | Key dependency |
|---|---|---|---|---|---|
| 1 | **National funding** | **Yes** | **High** | **Stable** | National economic situation |
| 2 | **Infrastructure model** | **Yes** | **High** | **Stable** | Research spending by funding agencies |
| 3 | **Institutional support** | **Yes** | **High** | **Stable** or Cyclic | Institutional funds availability |
| 4 | **Research project grants** | **Yes** | **High** | Cyclic - grants renew every 3–5 years | Infrastructure/research spending by funding agencies |
| 5 | **Content licensing/industrial support model** | No | Low | Function of usage | Commercial partner |
| 6 | **User subscription fees** | No | Low | Function of usage | Usage |
| 7 | **Value-added/asymmetrical pricing model (or freemium service)** | Not completely | Low | Function of usage | Usage |
| 8 | **Infrastructural razor & blades** | No | Low | Function of usage | Usage |
| 9 | **Public-private consortium** | **Yes** | **High** | **Potentially stable** | Commercial partner |
| 10 | **Online advertising & Corporate sponsorship** | **Yes** | **High** | Function of usage | Usage, commercial partners |
| 11 | **Open source volunteer (wiki approach)** | **Yes** | **High** | Highly dependent on participation | Willingness to contribute |
| 12 | **Donations** | **Yes** | **High** | **Potentially stable** | Partners |

ICM

The advantages of providing access to scientific data is evident , even for the provider, but psychological resistance is still well-established
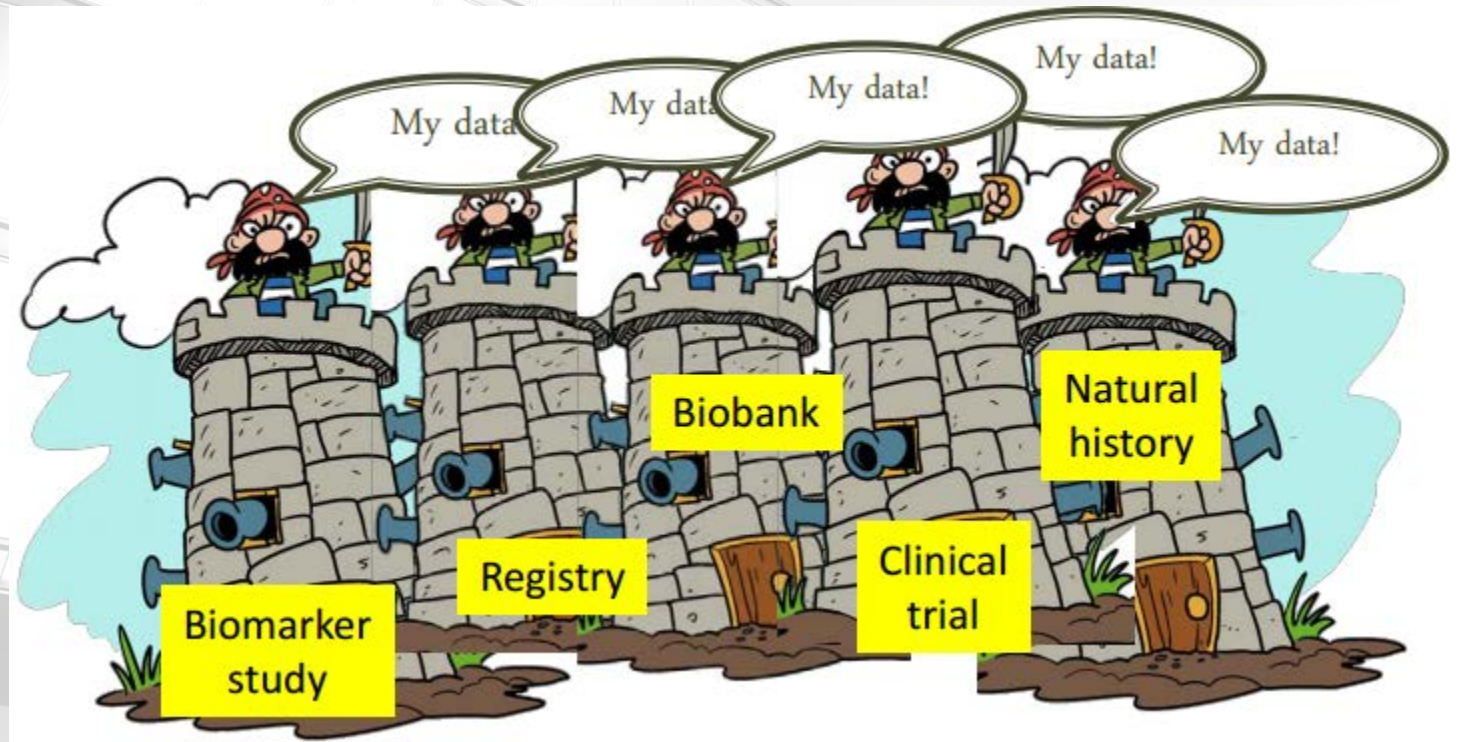
# Open Science does not mean Open Bar
# Open data means FAIR data

- FINDABLE
  - Means that the existence of data sets is public information

- ACCESSIBLE
  - Means that the rules to access the dataset are defined and public
  - From free access to restricted access to team members

- INTEROPERABLE
  - Means that the dataset respect the standards when available
  - Semantic standards (ontologies) + technical standards

- REUSABLE
  - Means that the dataset is annorated enough as to be re-used by third parties, and research results are publishable because the source is reliable

# FAIRification process
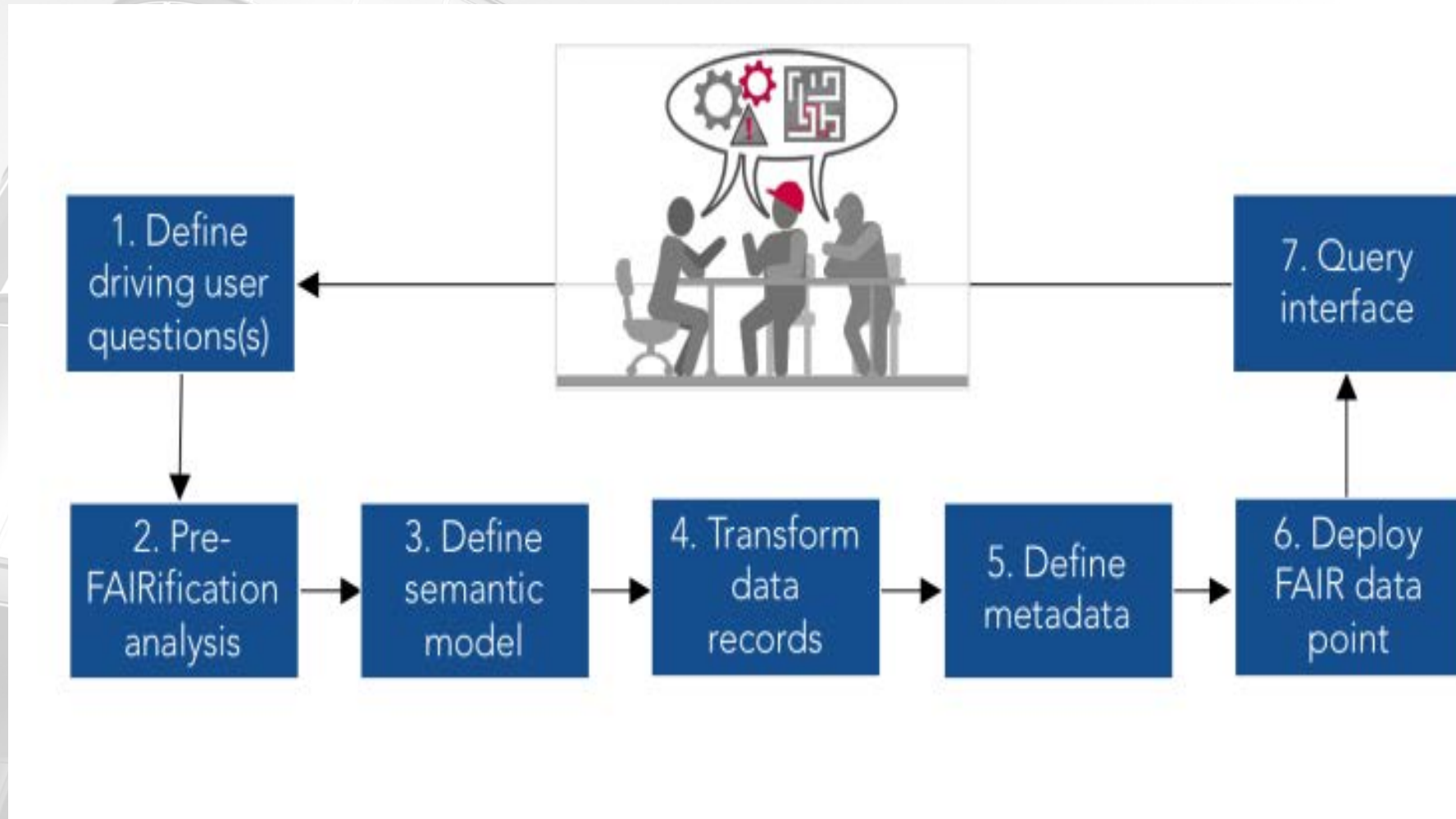
# ICM is on the way for a fairification of its data sets !

Thank you