The CWTS publication dataset

CWTS publication dataset

RISIS

- Initially only LeidenRanking set;
- CWTS version of Web of Science (SCI, SSCI, AHCI and CPCI);
- Main enhancements:
  - Citation matching
  - Author affiliations and funding organizations
  - CWTS LeidenRanking dataset
  - Publication-level classification
  - External info added.

Affiliations

RISIS 

- Continuous effort cleaning and harmonizing (desktop research)
- Permanent ID: OrgRegID and FirmRegID
- Link to other RISIS datasets (Patents, EUPRO, …)

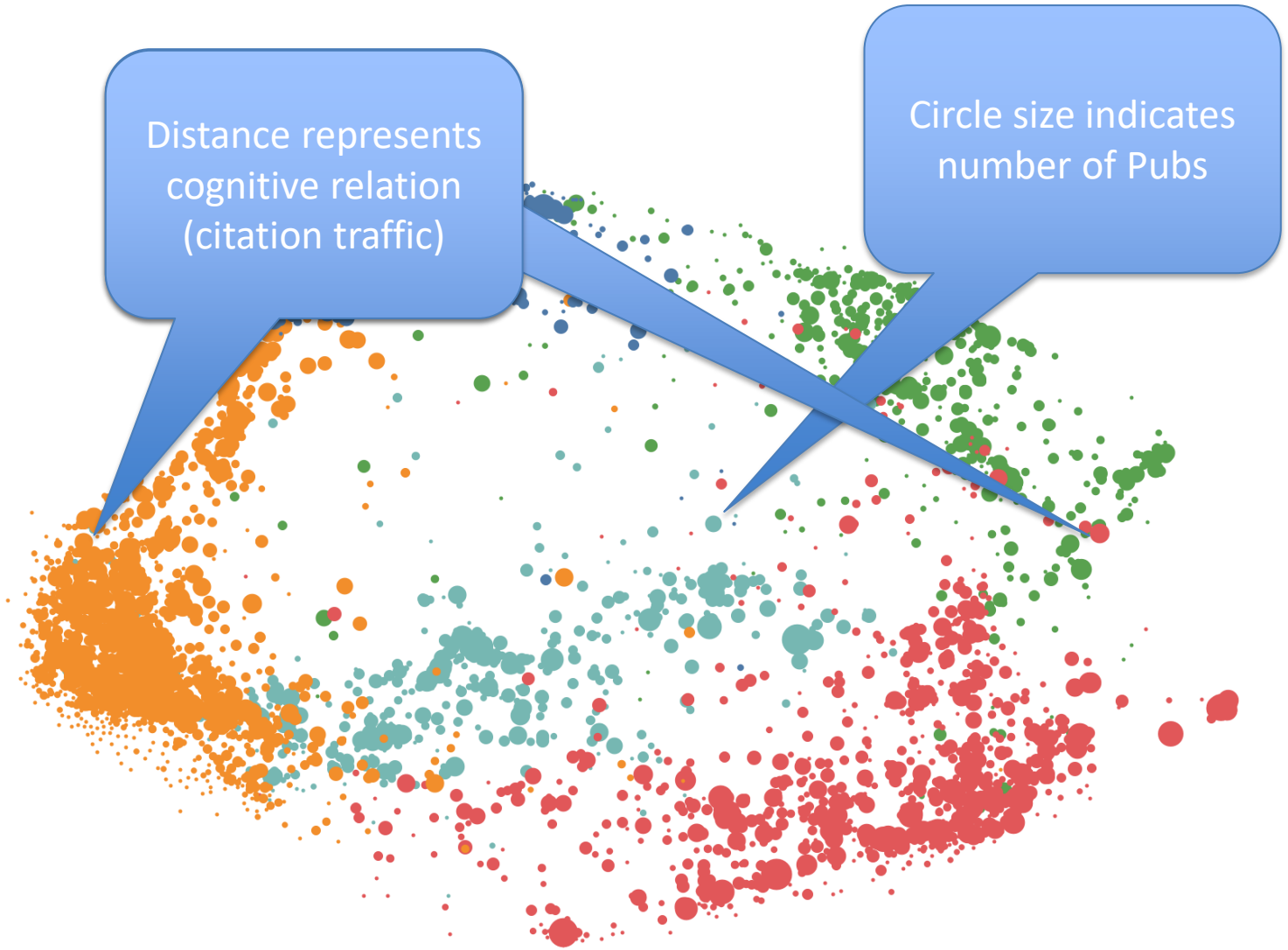
Publication level classification

RISIS

- Structure of all sciences (i.e., the CWTS dataset);
- Multiple level clustering of individual publications (~30 million);
- Independent from journal or journal classification;
- Objective structure based on 'input' researchers: a self-organizing structure;
- Challenging elements:
  - Label clusters
  - Updates.

Typical representation (~4000 clusters)

RISIS



Information per cluster/area/community

RISIS

- Number of pubs (full period and per year);
- Citation data (Avg cits per pub, etc);
- Author statistics (Avg number of authors);
- Other characterizing indicators
  - Proportion of OA publications
  - Proportion of papers with acknowledgement to (a specific) funder
  - Proportion International collaboration papers
  - Proportion papers with industry/ hospital/ … involved
  - Proportion of papers not in English.

And using external data linked

RISIS

- Proportion of papers mentioned on Twitter
- Proportion of papers mentioned in news items
- Proportion of papers mentioned in policy documents
- …

Area-based connectedness (ABC): communities relating to society

Area-based Connectedness (ABC)

RISIS

- R&D relevance to society and vv.;
- Such relevance measured to the community level (rather than at the actor level);
- Characterizing communities (publication clusters), related to their connectedness;
- Enriching publication output by the character of their community (a publication inherits the character of the cluster it belongs to);
- The stronger the connectedness, the higher the relevance.

Key assumptions

RISIS

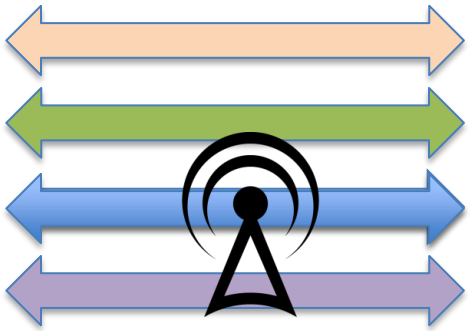
- Societal impact too *diverse* and *complicated* to assess in a 'traditional' quantitative way;
- *Societal connectedness* is a more productive approach;
- *Connectedness* (like *societal impact*) is not a merit of one actor only. It is a credit of a community/ research area.

Which one is more important?

Connectedness

How to measure connectedness of research?

RISIS



- ***Academic output*** connected to ***society;***
- Through traces in **output**/ signals between ***research output*** and ***society;***
- Each trace/ signal represents a certain link/ connection/ interaction,
  a ***dimension of connectedness.***

Traces/ signals and dimensions

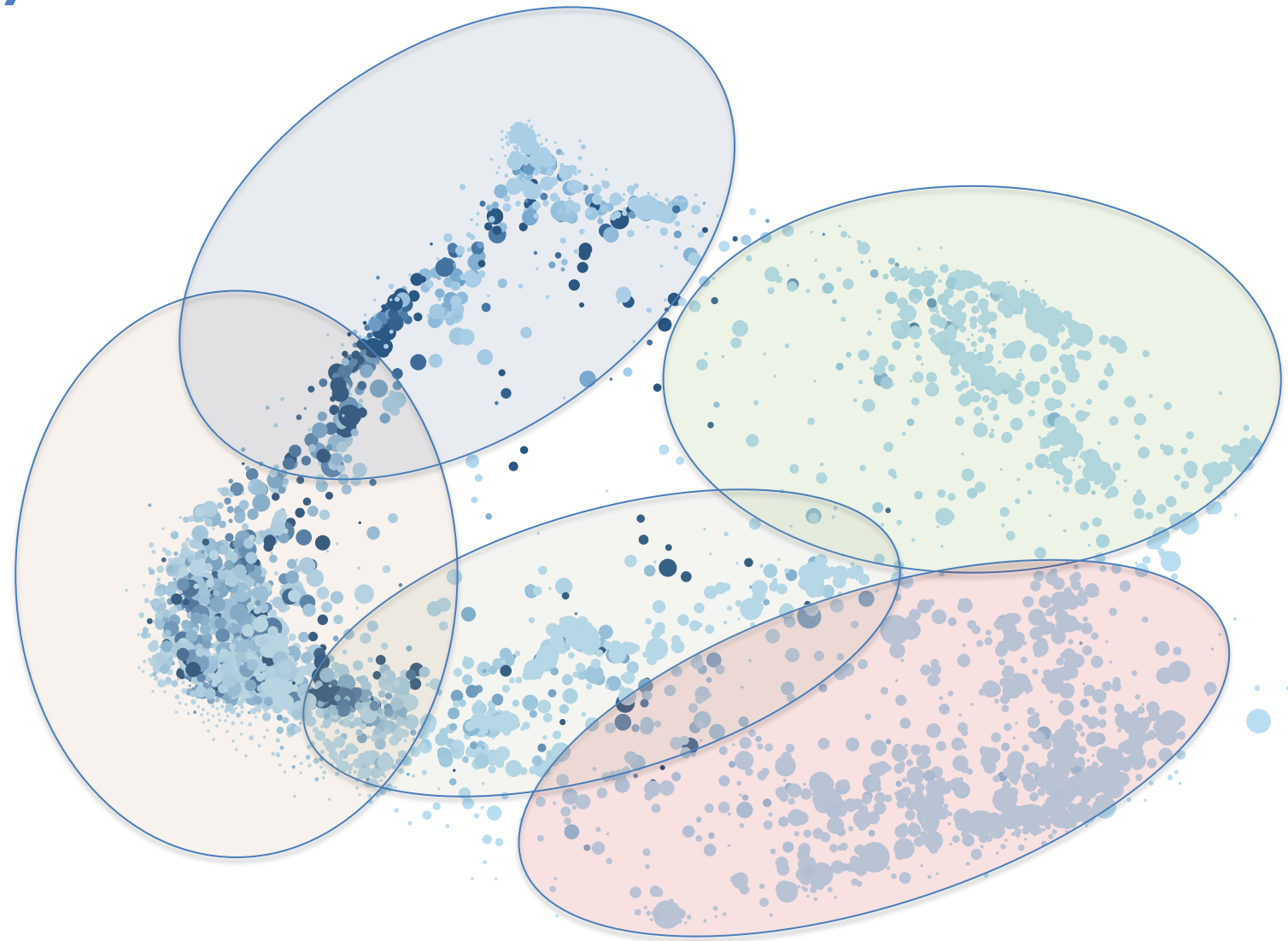
RISIS

Signal	Dimension
Papers (co-)authored by industry	Relevance Industry to R&D and vv.
Papers (co-)authored by non-academic hospitals	Relevance local hospitals to R&D and vv.
Papers published in local languages	Relevance R&D for local audience and vv.
Papers cited by patents	Relevance R&D for Technological development or market
Papers mentioned on twitter (or other social media)	Relevance R&D for general public and vv.
Papers mentioned in policy documents	Relevance R&D for policy
Papers mentioned in news	Relevance R&D for general public and vv.
Papers funded by EC	Relevance EC for R&D and vv.

Distribution of traces/ signals across all science

Policy

RISIS

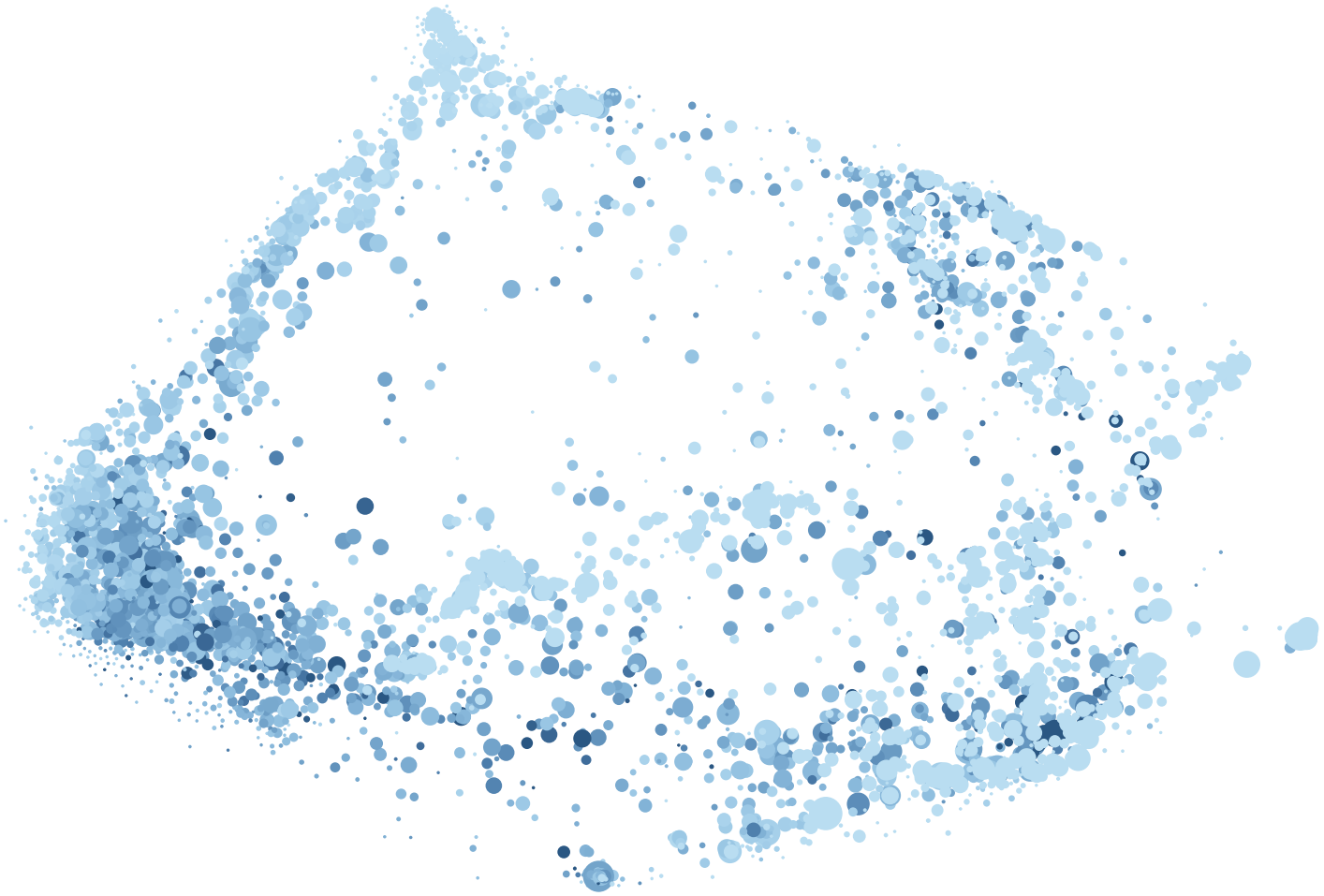


Research primarily connected to policy:

- Social sciences
- Cognitive psychology
- Clinical studies
- Life & earth

Industry

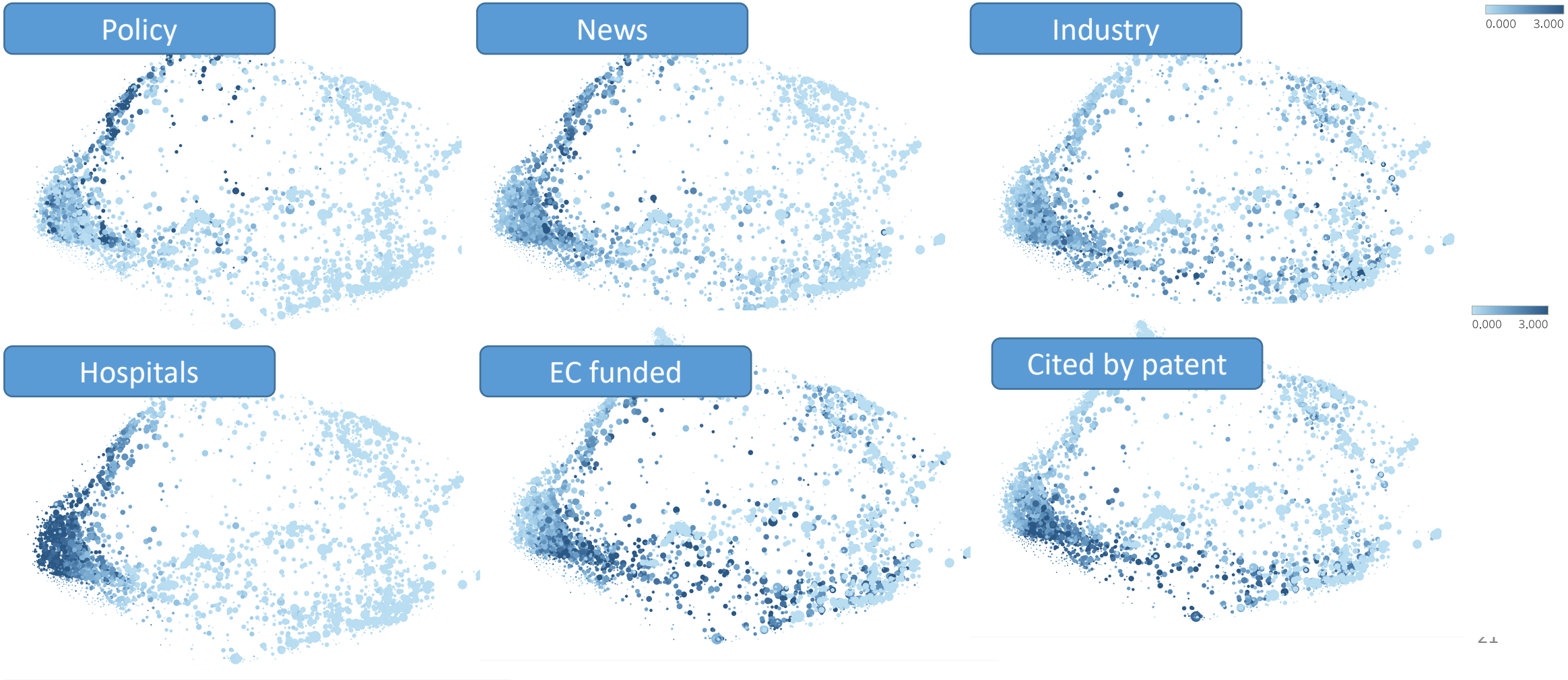
RISIS



Research primarily connected to industry:

- Biomedical/ pharmacy
- Natural sciences & engineering
- Computer science

ABC distribution across all science



Which opportunities does this create?

An example:
EC funding through the "cancer lens"

RISIS

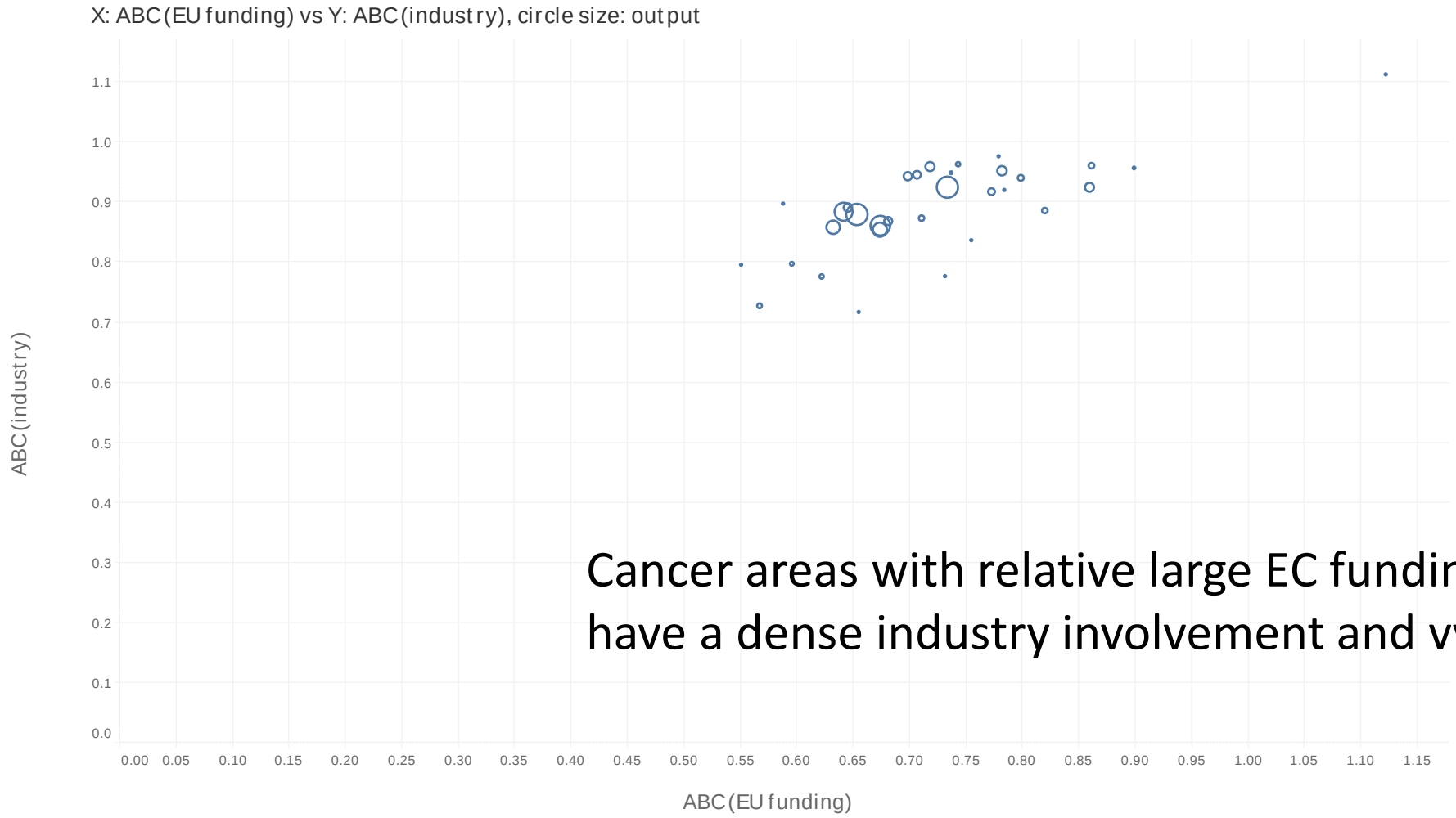
- Cancer publications by country;
- Distribution across communities/ areas;
- Communities/ areas characterized by (ABC)
  - EC funding
  - Industry involvement
  - Local Hospital involvement
  - Local interest.

Result

RISIS 

- A Country's cancer research output (publications);
- Characterized by/ via the areas in which it is active:
  - EU funding
  - Industry involvement
  - Local hospital involvement
  - Local interest;
- Country-wise correlation between characterizations.

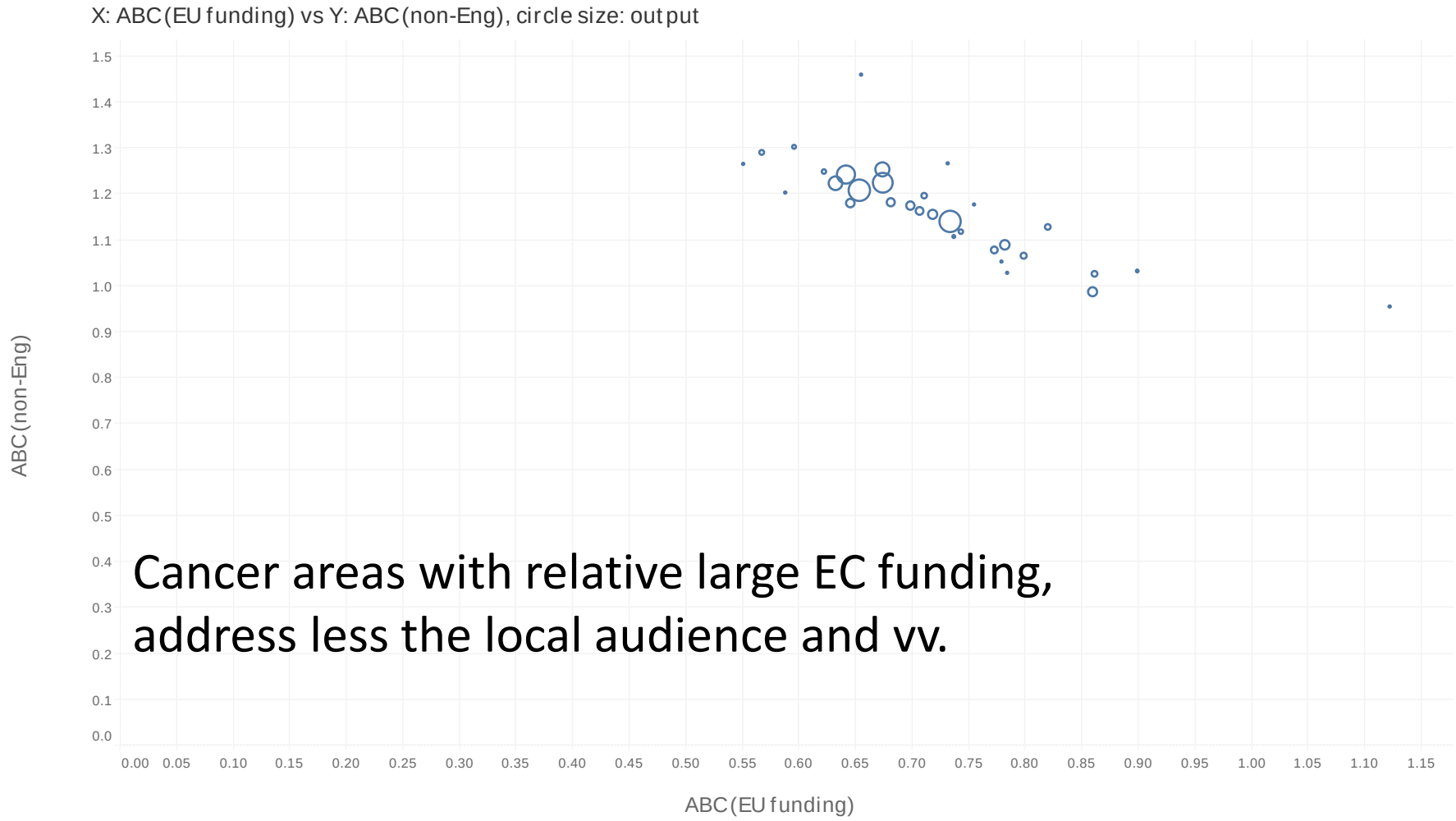
EC funding vs Industry involvement
(by European country)



Cancer areas with relative large EC funding, also have a dense industry involvement and vv.

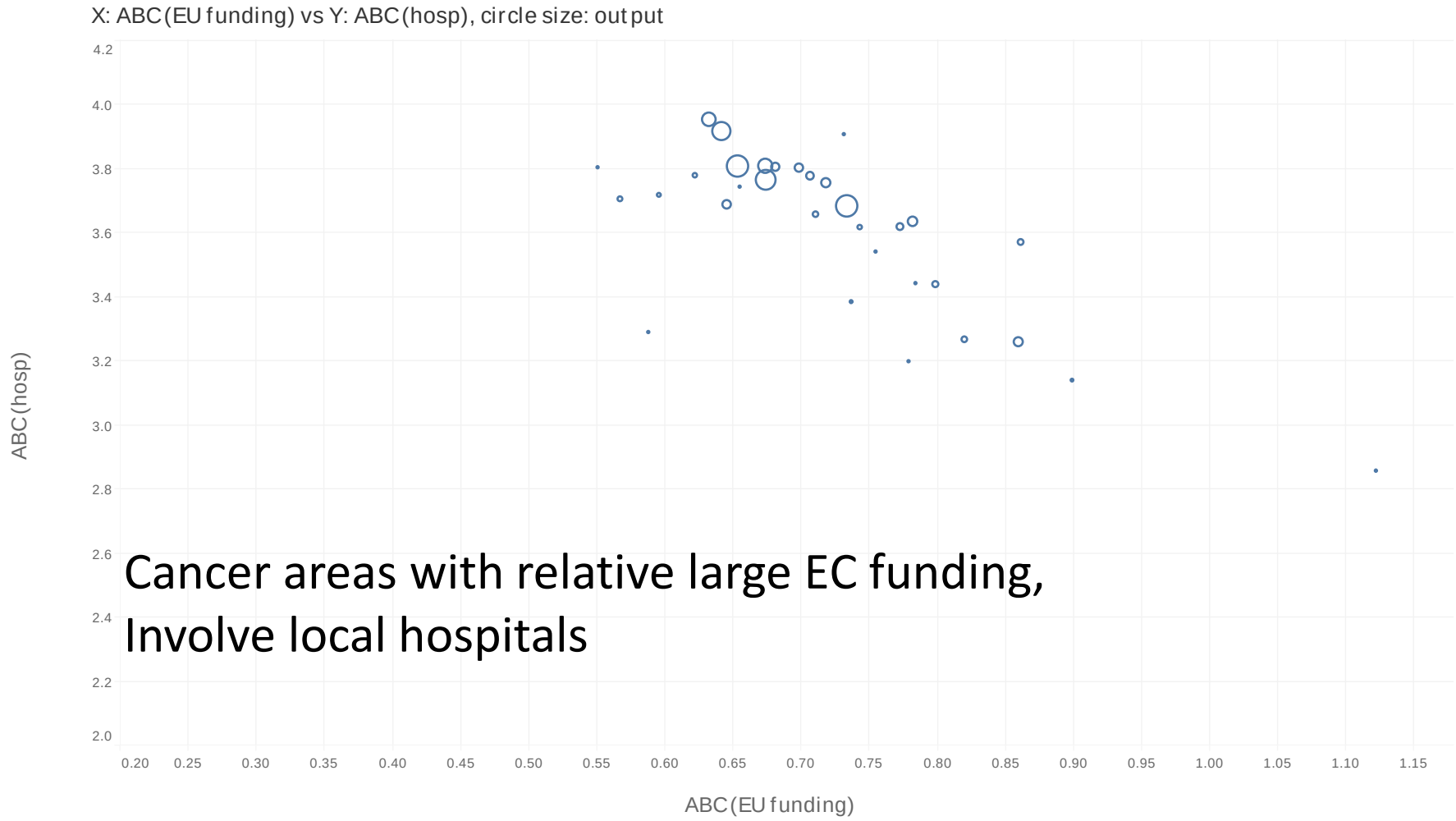
EC funding vs local interest
(by European country)

RISIS



Cancer areas with relative large EC funding, address less the local audience and vv.

EC funding vs Local hospital involvement
(by European country)



Cancer areas with relative large EC funding,
Involve local hospitals

Insights

RISIS 

- Community/area-based characterizations provide new insights and a world of opportunities;
- Adding information to publication output provides such characterizations;
- This provides potential links to societal challenges/ missions/ SDGs;

- The results discussed seem to point to:
  - EC funding in cancer research involves industry, less hospitals, how does this relate to co-creation?
  - EC funding in cancer research involves less the local audience, how does this relate to smart specialization?

Related literature and references

RISIS

On portfolios:

- Sarewitz, D., & Pielke Jr, R. A. (2007). The neglected heart of science policy: reconciling supply of and demand for science. environmental science & policy, 10(1), 5-16.

On plural and conditional advice:

- Stirling, A. (2010). Keep it complex. Nature, 468(7327), 1029-1031.

On priority setting

- Yegros-Yegros, A., van de Klippe, W., Abad-Garcia, M.F. et al. Exploring why global health needs are unmet by research efforts: the potential influences of geography, industry and publication incentives. Health Res Policy Sys 18, 47 (2020). https://doi.org/10.1186/s12961-020-00560-6
- Rafols, I; Yegros, A (2017) Is research responding to health needs? [Blog post] Retrieved from https://observatoriosociallacaixa.org/en/-/responde-la-investigacion-a-las-necesidades-de-salud

On community based approach

- Noyons, E. (2019). Measuring Societal Impact Is as Complex as ABC. Journal of Data and Information Science, 4(3), 6–21. https://doi.org/10.2478/jdis-2019-0012

RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

THANK YOU !

WWW.RISIS2.EU

ZENODO.ORG/COMMUNITIES/RISIS

CONTACT@RISIS2.EU

@RISIS_EU

FACEBOOK.COM/RISIS.EU

RISIS2 EU PROJECT

Université Gustave Eiffel · AIT Austrian Institute of Technology · Politecnico Milano 1863 · Consiglio Nazionale delle Ricerche · Universiteit Leiden · The University Of Sheffield. · Fraunhofer ISI · Università della Svizzera Italiana · JOANNEUM RESEARCH

University of Sussex · DZHW Deutsches Zentrum für Hochschul- und Wissenschaftsforschung · University of Strathclyde Glasgow · Sapienza Università di Roma · NIFU Nordisk institutt for studier av innovasjon, forskning og utdanning · CSIC Consejo Superior de Investigaciones Científicas · Samuel Neaman Institute for National Policy Research · Technology Centre ASCR · ATHENA Research & Innovation Information Technologies