
# Otvoreni pristup istraživačkim podacima

Alen Vodopijevec [alen@irb.hr] / Institut Ruđer Bošković, Zagreb

# Plan za danas

- **Zašto** istraživački podaci (IP) trebaju biti javno dostupni?
  - općenito o podacima
- **Kako** istraživačke podatke učiniti javno dostupnima?
  - **Planiranje** upravljanja istraživačkim podacima
    - organizacija i dokumentiranje podataka
    - pohrana i sigurnost
    - intelektualno vlasništvo i etička pitanja
    - dijeljenje podataka i pravo na pristup

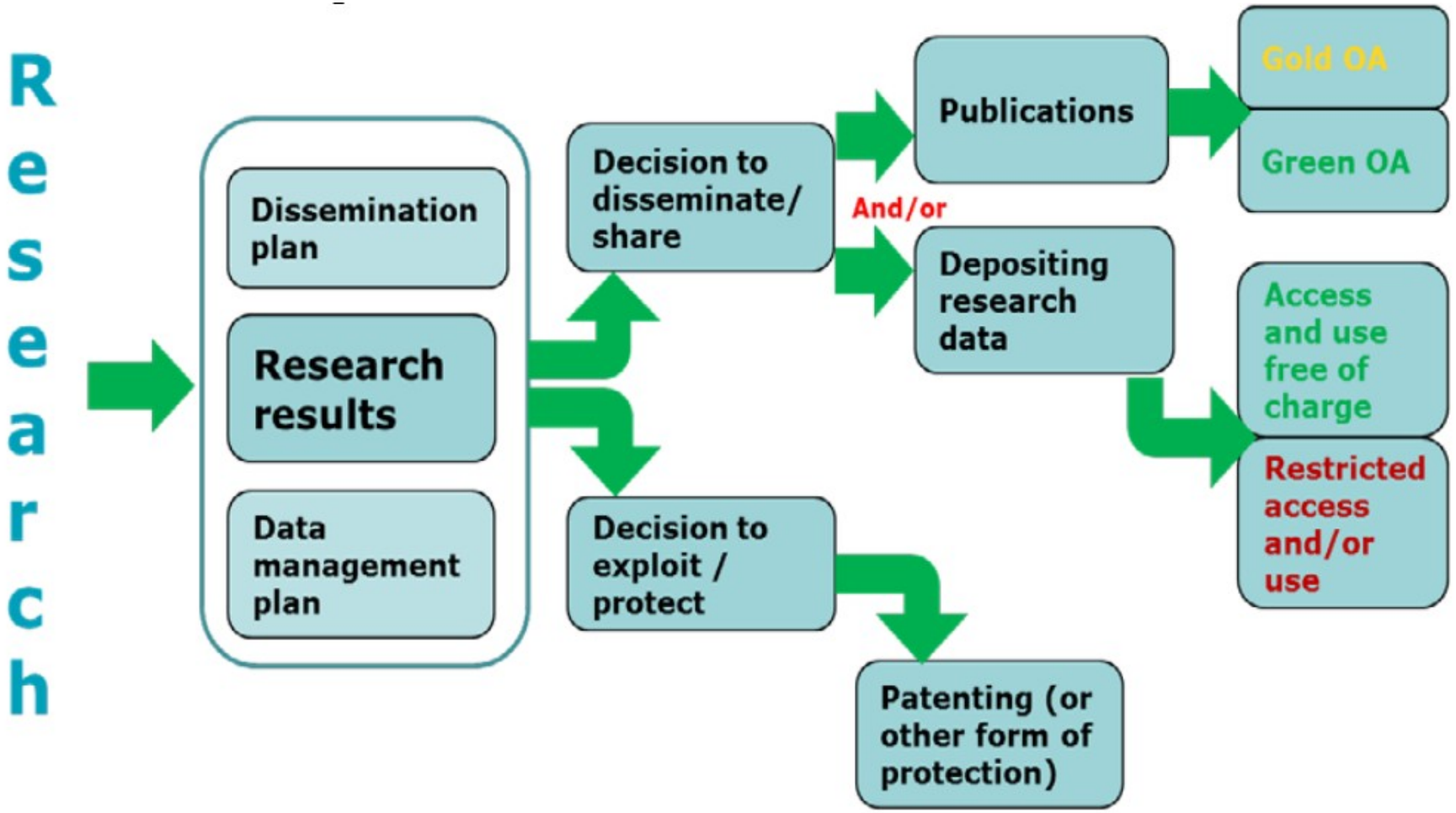
FOSTER

# Imperativ dostupnosti podataka

- Znanost počiva na akumulaciji znanja (i promjenama paradigmi)
  - Napredak temeljen na postojećim spoznajama
    - Dijeljenje i ponovno korištenje
    - Suradnja
- Da bi se osiguralo **neometano funkcioniranje i napredak znanosti** izvorni (primarni) **istraživački podaci moraju biti slobodno dostupni**!*

FOSTER

# Otvoreni pristup




FOSTER

# H2020 Open Data Pilot (1)

- Obveza pohrane i objave istraživačkih podataka nastalih u okviru projekata financiranih sredstvima EK (javna sredstva)

- Smjernice za otvoreni pristup publikacijama i podacima u okviru H2020
  http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

FOSTER

# H2020 Open Data Pilot (2)

- Izrada Plana za upravljanje istraživačkim podacima tijekom prvih 6 mjeseci trajanja projekta
- Mogućnost izuzeća od te obveze (opt out)
- Objava izvornih podataka potrebnih za validaciju usporedo s objavom rada
- Omogućavanje besplatnog pristupa podacima korištenjem CC-BY ili CC0 licenci

FOSTER

# H2020 Open Data Pilot (3)

**29.3 Open access to research data**

*[**OPTION for actions participating in the open Research Data Pilot**: Regarding the digital research data generated in the action ('**data**'), the beneficiaries must:*

*(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:*

*(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;*

*(ii) other data, including associated metadata, as specified and within the deadlines laid down in the '**data management plan**' (see Annex 1);*

*(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).*

*This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.*

*As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex 1, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.]*

FOSTER

Article 29.3 of the Model Grant Agreement

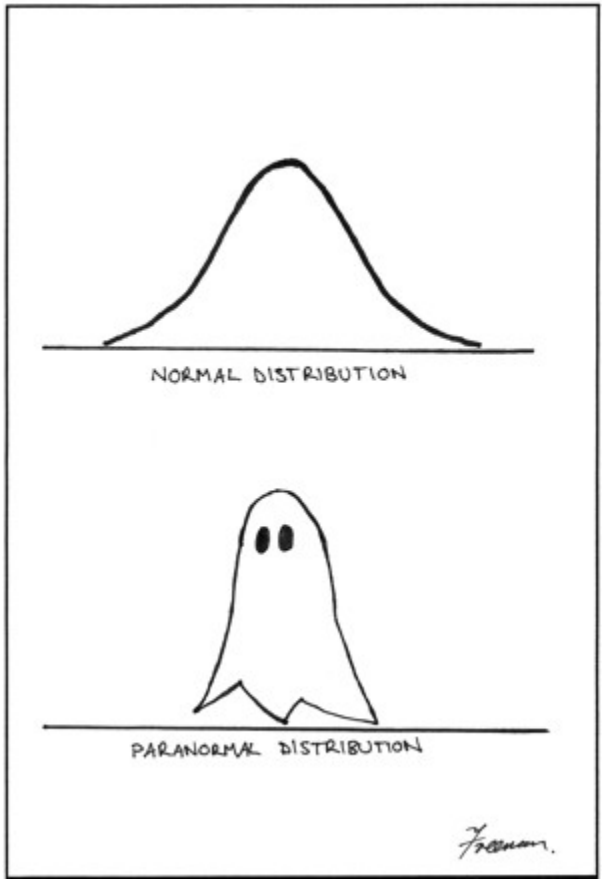
# Zašto dijeliti podatake?

- Prije interneta dostupnost podataka bila je ograničena i vremenskom i prostornom dimenzijom
  - Osim moralne
- Danas postoje svi preduvjeti za neometanim dijeljenjem podataka no još uvijek ne postoji spoznaja o korisnosti za cijelu zajednicu
  - Ostaje sporna moralna komponenta

FOSTER

# A zašto ne dijeliti podatke?

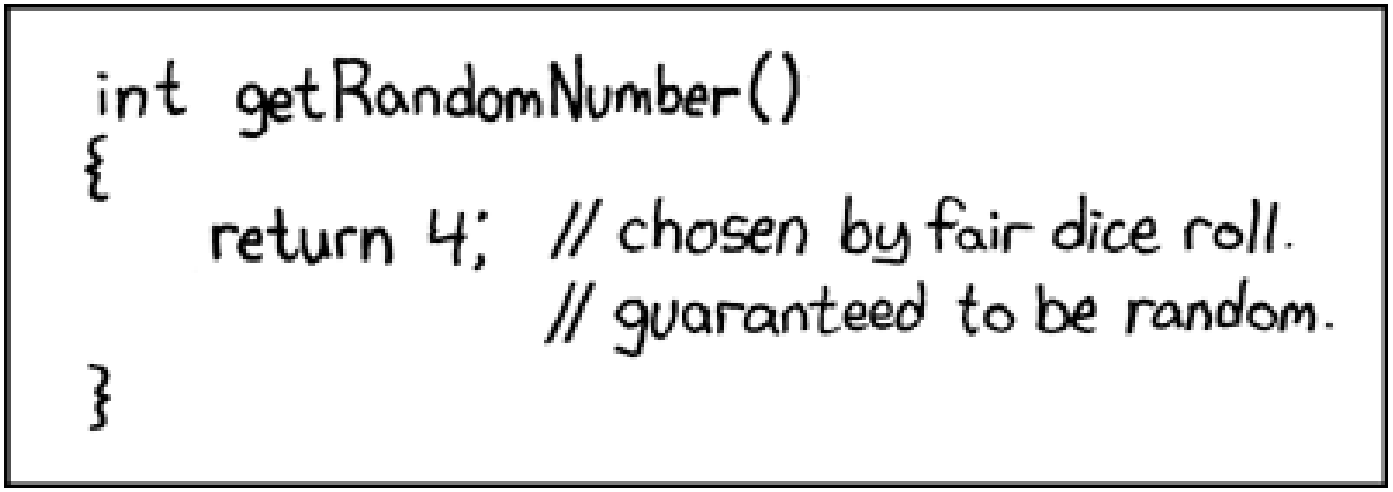
- Validacija rezultata
- Sekundarne analize
  - Meta analize
- Suradnja (GENOM)
- Obrazovanje




*A visual comparison of normal and paranormal distributions,* Matthew Freeman, J Epidemiol Community Health 2006;60:6. Lower caption says 'Paranormal Distribution' - no idea why the graphical artifact is occuring

FOSTER

# Prednosti - VALIDACIJA REZULTATA




http://xkcd.com/221/

- http://en.wikipedia.org/wiki/Scientific_misconduct
- Razne vrste 'znanstvenih nepodopština'*

FOSTER

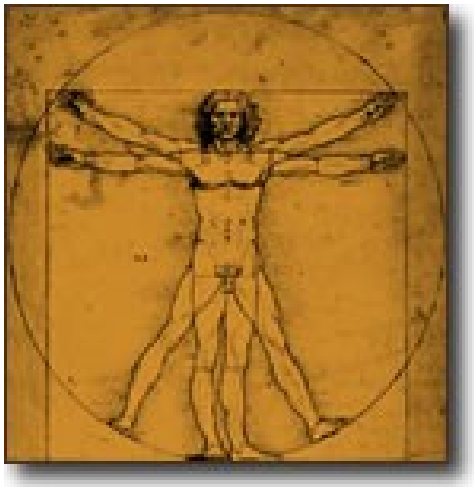
# Prednosti - SEKUNDARNE ANALIZE

- Ponovno korištenje postojećih podataka
  - Na drugačiji način
  - Testiranje drugačijih hipoteza
  - Ponavljanje istraživanja te usporedba s novoprikupljenih podataka
  - Meta analize
    - Usporedba s drugim izvorima podataka (?)

FOSTER

# Prednosti - SURADNJA

- GENOME projekt



http://www.genome.gov/10001772
http://unlockinglifescode.org/timeline?tid=4
ftp://ftp.ensembl.org/pub/release-77/fasta/homo_sapiens/dna/

FOSTER

# Prednosti - OBRAZOVANJE

FOSTER

# RESEARCHER DATA SHARING INSIGHTS

WILEY

- Wiley's Researcher Data Insights Survey was launched earlier this year to understand how and why researchers make their research data publicly available. The study's results, highlighted below, are intended to advance the global conversation about data sharing and help Wiley better meet the needs of our researchers, authors, and partners in the rapidly evolving landscape of scientific research and communications.
- The survey was deployed in March 2014 and received more than 2,250 responses from researchers around the world.

## GLOBAL DATA SHARING TRENDS

Data sharing practices vary widely across research fields and geographic areas. Just over half of researchers report making their data publicly available, though archiving results in repositories is not yet the norm.


### WAYS DATA IS SHARED

- **67%** As supplementary material in a journal
- **37%** Personal, institutional or project webpage
- **26%** Institutional data repository *(i.e. university or institute-sponsored)*
- **19%** Discipline-specific data repository
- **6%** General-purpose data repository *(e.g. Dryad, figshare)*
- ✓ **5%** Other

Globally, researchers also report sharing their data in limited and non-permanent ways: 57% are sharing data at a conference while 42% of researchers share their data upon informal request (e.g. email, direct contact, etc.).

## RESEARCHER MOTIVATIONS FOR SHARING DATA


## DATA SHARING TRENDS BY COUNTRY

**UNITED STATES**
**46%** SHARING
**54%** NOT SHARING

Among researchers in the US sharing their data publicly, two out of three do so because it is standard practice in the communities and because they believe it benefits the public. Similar to their counterparts in the UK, the majority of US-based researchers also share data to increase the impact or visibility of their research.

**UNITED KINGDOM**
**43%** SHARING
**57%** NOT SHARING

While more than 40% of UK researchers are sharing data, only about 14% are using discipline-specific or other public repositories like Dryad and figshare. The two key drivers that motivate UK researchers to share their data are the prospect of gaining increased impact or visibility for their work and to satisfy funder requirements.

**JAPAN**
**44%** SHARING
**56%** NOT SHARING

Compared with their counterparts around the world, researchers in Japan cite concerns about being scooped as a reason for not sharing data more frequently. Nearly five out of ten Japanese researchers point to this as a reason for not sharing their data, roughly double the global average.

**CHINA**
**36%** SHARING
**64%** NOT SHARING

Nearly five in ten Chinese researchers say they are not sharing data because they are not required to do so by their funders or institutions. They are more likely than their global counterparts to say that they do not see data sharing as a personal responsibility and plan to take direction from funders to guide their data sharing decisions in the future.

**BRAZIL**
**52%** SHARING
**48%** NOT SHARING

Two out of three researchers in Brazil say that a guarantee of proper credit or attribution would compel them to share more of their data publicly in the future.

**AUSTRALIA**
**41%** SHARING
**59%** NOT SHARING

Researchers in Australia say they would be most incentivized to make their data accessible in the future to ensure preservation as well as transparency and re-use. The majority of researchers also ranked funder requirements among top reasons to share in the future.

**GERMANY**
**55%** SHARING
**45%** NOT SHARING

Among German researchers sharing their data publicly, three out of four are driven to share data because they believe it will increase the visibility of their research and want to ensure public transparency and re-use. About 20% of German researchers are making use of general purpose repositories (like figshare and Dryad), significantly more than their counterparts around the world, including those in the US and UK.

## REASONS WHY RESEARCHERS ARE HESITANT TO SHARE THEIR DATA

- **42%** Intellectual property or confidentiality issues
- **36%** My funder/institution does not require data sharing
- **26%** I am concerned that my research will be scooped
- **26%** I am concerned about misinterpretation or misuse
- **23%** Ethical concerns
- **22%** I am concerned about being given proper citation credit or attribution
- **21%** I did not know where to share my data
- **20%** Insufficient time and/or resources
- **16%** I did not know how to share my data
- **12%** I don't think it is my responsibility
- **12%** I did not consider the data to be relevant
- **11%** Lack of funding
- **7%** Other

## DATA SHARING BY DISCIPLINE

Data sharing, specifically by way of data repositories, is most prevalent amongst life scientists, particularly those in the earth and environmental and agriculture and food sciences.


**Where Health Scientists share their work:**

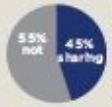
- **68%** As supplementary material in a journal
- **29%** Personal/institutional/lab webpages
- **29%** Institutional data repositories *(i.e. university or institute-sponsored)*
- **27%** Discipline-specific data repositories
- **5%** General-purpose data repositories *(e.g. Dryad, figshare)*

A typical *Health Science researcher* says she would be motivated to share her data in the future in order to benefit the public, so long as privacy and ethical concerns are assuaged.


**Where Life Scientists share their work:**

- **76%** As supplementary material in a journal
- **42%** Discipline-specific data repositories
- **29%** Personal/institutional/lab webpages
- **25%** Institutional data repositories *(i.e. university or institute-sponsored)*
- **13%** General-purpose data repositories *(e.g. Dryad figshare)*

A typical *Life Science researcher* says she would be motivated to share more of her data in the future if she was guaranteed proper credit.




**Where Physical Scientists share their work:**

- **60%** As supplementary material in a journal
- **41%** Personal/institutional/lab webpages
- **28%** Institutional data repositories *(i.e. university or institute-sponsored)*
- **10%** Discipline-specific data repositories
- **3%** General-purpose data repositories *(e.g. Dryad figshare)*

A typical *Physical Science researcher* says she would be motivated to share her data in the future because it is standard practice within her research community and because it increases the impact and visibility of her work.


**Where Social Scientists share their work:**

- **52%** As supplementary material in a journal
- **51%** Personal/institutional/lab webpages
- **25%** Institutional data repositories *(i.e. university or institute-sponsored)*
- **3%** General-purpose data repositories *(e.g. Dryad, figshare)*
- **2%** Discipline-specific data repositories

A typical *Social Science and Humanities researcher* says she would be motivated to share her data in the future if it increased the impact and visibility of her work or if she was required to by her funder.

FOSTER

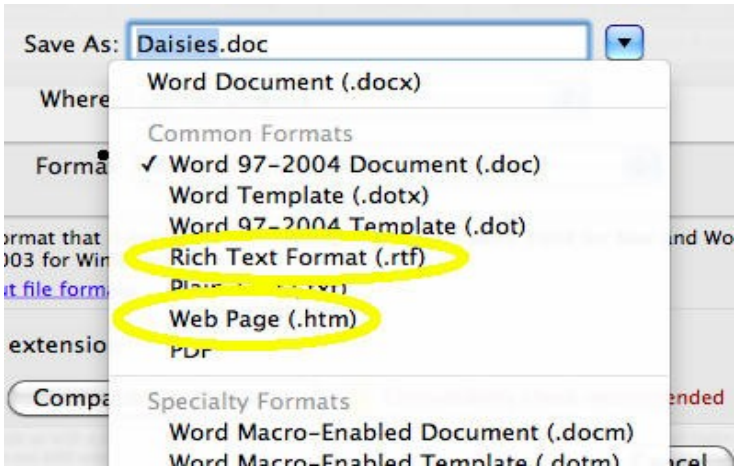
Izvor: http://exchanges.wiley.com/ (20.11.2014.)

# ULOGA ARHIVIRANJA
# U PROCESU OTVARANJA PRISTUPA PODACIMA

FOSTER

# Format podataka

- kad govorimo o izvornim istraživačkim podacima podrazumijevamo neki od digitalnih formata
- izvorno digitalno ili digitalizirano


FOSTER

# Niz različitih podjela

Podaci mogu biti:

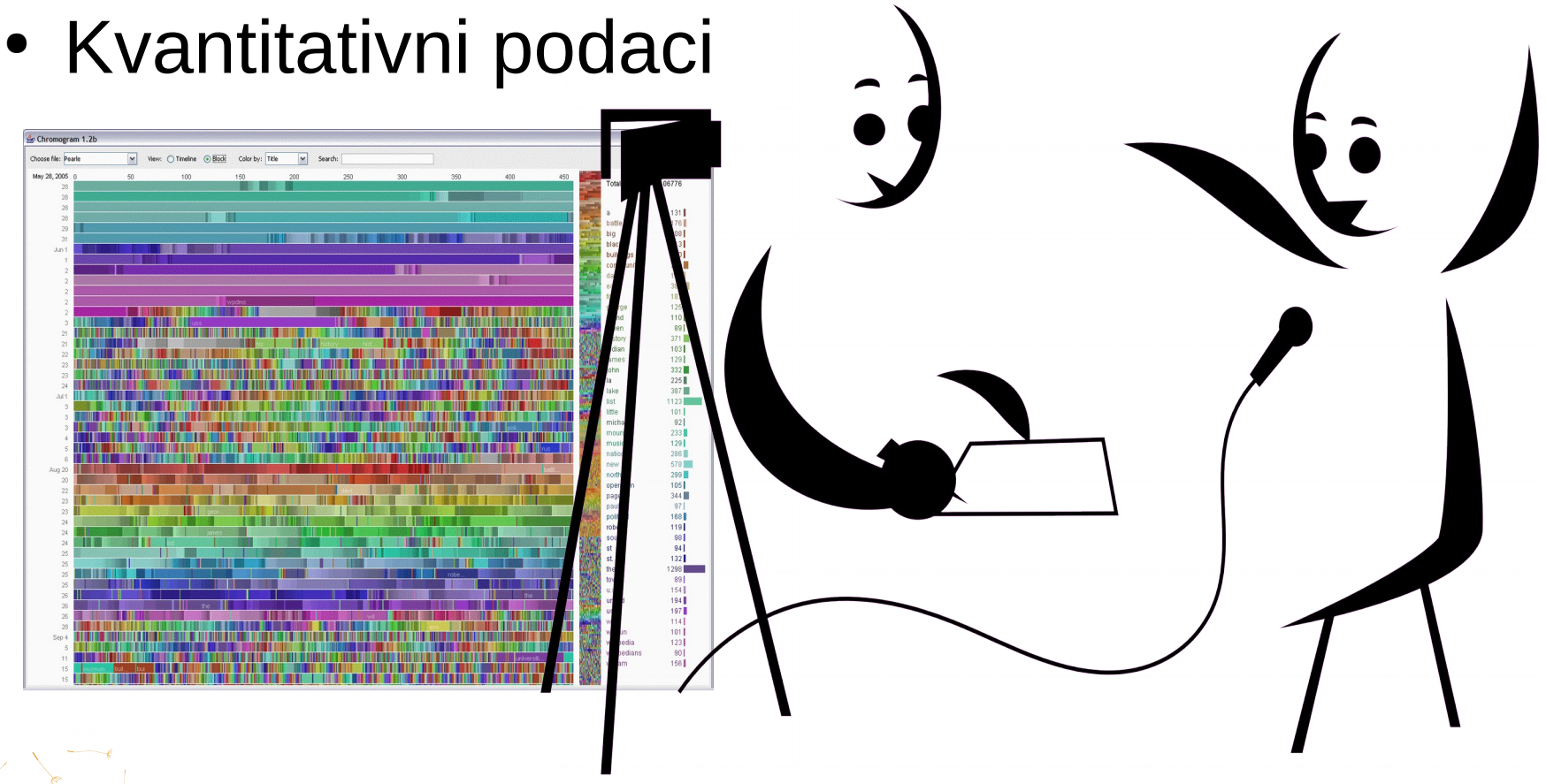
- tekstualni / numerički / multimedijalni
- modeli / softver
- disciplinarno specifični
- karakteristični za određeni instrument

*Podjele s obzirom na različite kriterije(?)*

FOSTER

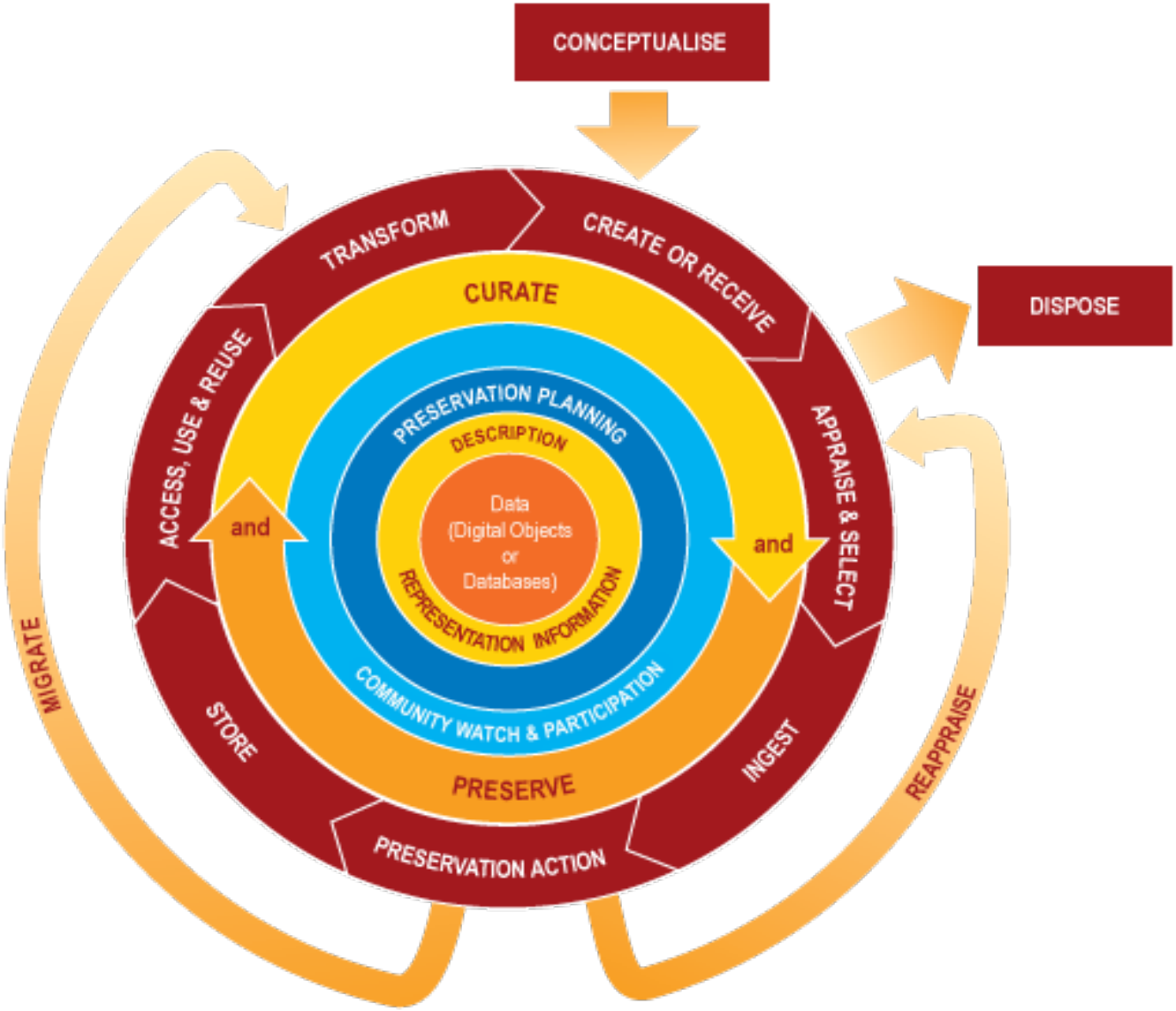
# Osnovna podjela istraživačkih podataka

- Kvalitativni podaci
- Kvantitativni podaci



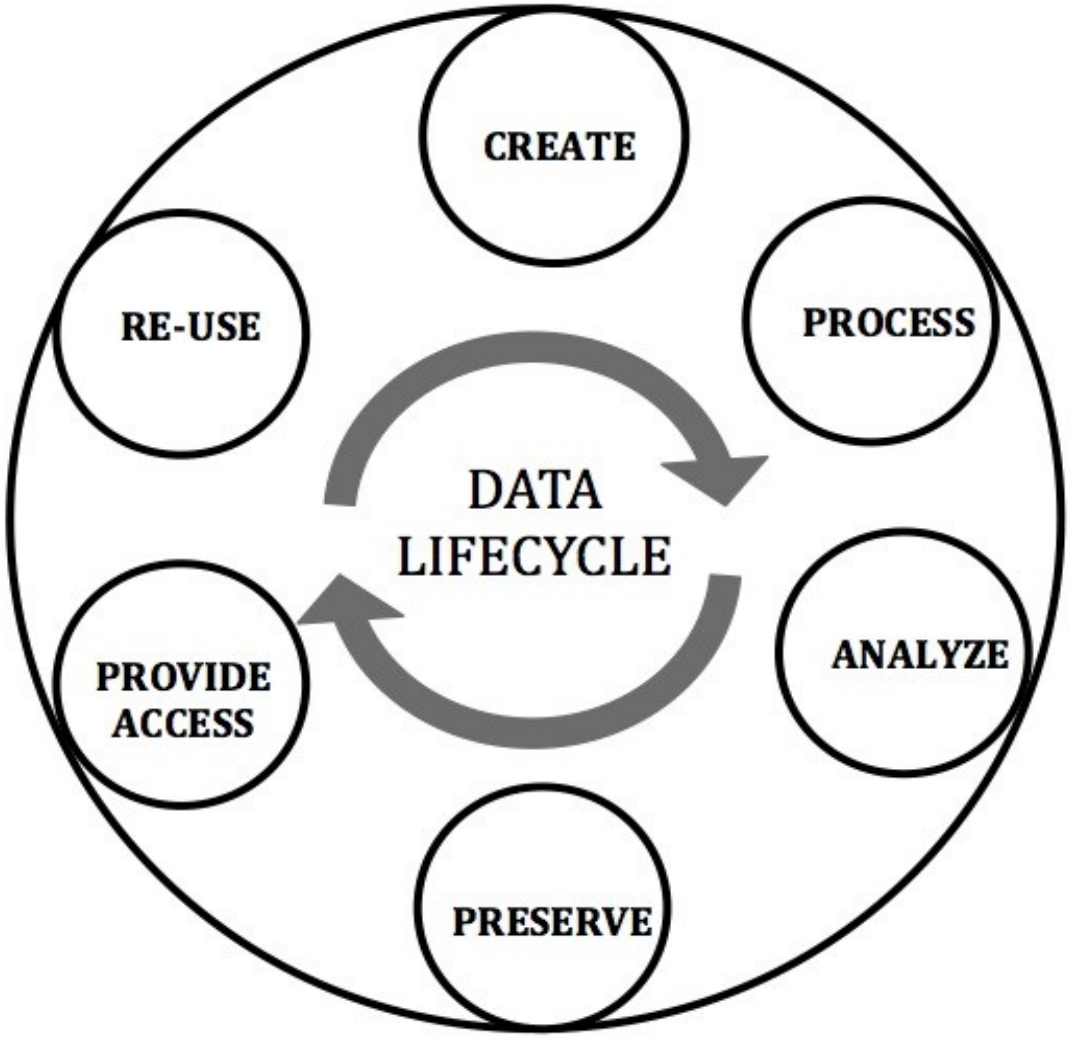

FOSTER

# Arhiviranje podataka




FOSTER

# Pojednostavljeni životni ciklus



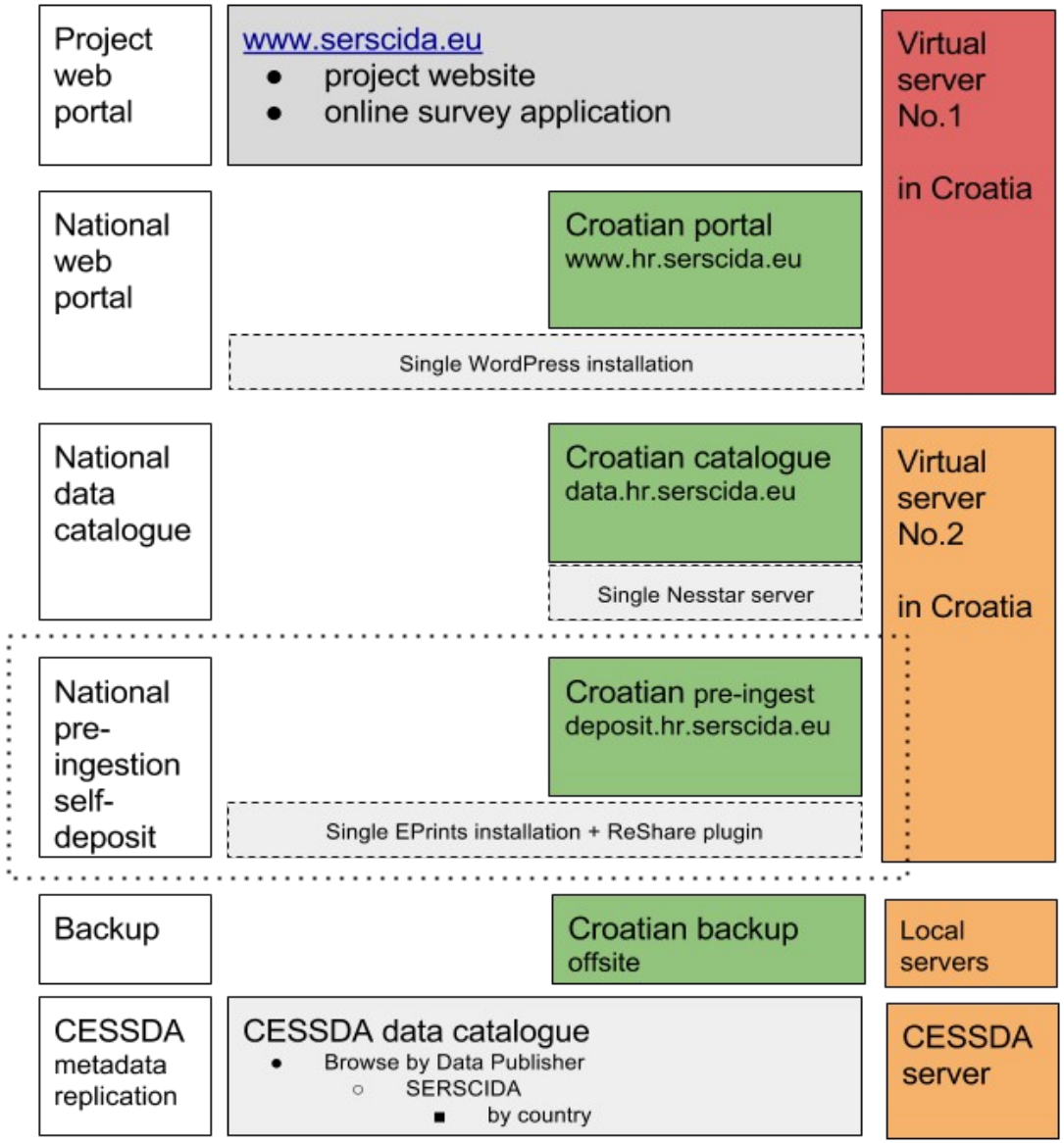

FOSTER

# Životni ciklus podataka

- **Generiranje** podataka - osmišljavanje istraživanja, plana upravljanja istraživačkim podacima, obrasci pristanka na dijeljenje, identifikacija postojećih dostupnih podataka, prikupljanje (eksperiment, opažanje, mjerenje, simulacija)
- **Procesiranje** podataka – priprema za analizu
- **Analiza** podataka
- **Prezervacija** podataka - pohrana i sigurnose kopije, arhiviranje, opisivanje metapodacima, migracija u formate prikladne za dugotrajno očuvanje
- **Osiguravanje pristupa** podacima - dijeljenje podataka, kontrola pristupa
- **Ponovno korištenje** podataka – re-use

FOSTER

# Primjer: SERSCIDA [www.hr.serscida.eu]




FOSTER

# KAKO i GDJE OBJAVITI PODATKE?

## Stvaranje plana upravljanja istraživačkim podacima

FOSTER

- Kada ne postoji specijalizirani arhiv?
- Razmisliti o:
  - Vidljivosti i dostupnosti pohranjenih podataka
  - Reguliranju prava pristupa
  - Razumljivosti i pristupačnosti
  - Prezervaciji

FOSTER

# Objava podataka - VIDLJIVOST

- Kako će potencijalni korisnici pronaći konkretne podatke?
  - Institucijski repozitorij
  - ZENODO (OpenAIRE projekt)
  - Katalozi tematskih repozitorija
    - Databib
    - Re3data.org
  - Repozitorij časopisa(?)
- Trajni/postojani identifikatori(?)

FOSTER

# Objava podataka - PRAVA PRISTUPA i PRAVILA KORIŠTENJA

- **Tko**, **kada** i **pod kojim uvjetima** može **pristupiti** podacima te **na koji način** ih smije **koristiti**?
- Panton Principles
  - Korištenje što otvorenijih licenci
  - http://pantonprinciples.org/
- Definicija otvorenog znanja I usporedba licenci, [infografika]
- Pravna i etička pitanja!

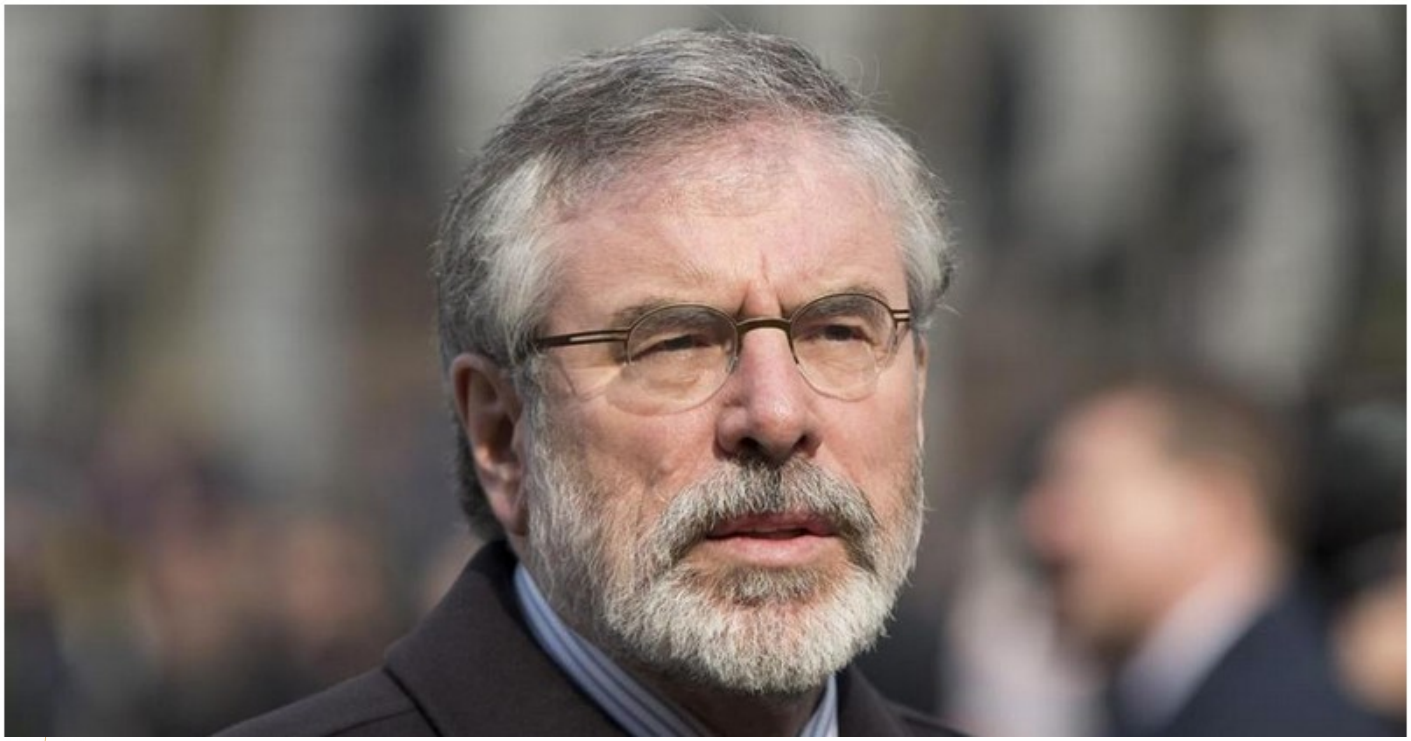
FOSTER

# Primjer (1)

## BC's Irish project leads to arrest of Gerry Adams

By Mark Arsenault, Zachary T. Sampson and Jeremy C. Fox | GLOBE STAFF AND GLOBE CORRESPONDENTS MAY 01, 2014

51 COMMENTS



FOSTER

# Objava podataka - RAZUMLJIVOST i PRISTUPAČNOST

- Kako i na koji način će podaci biti opisani
  - Dokumentacija
  - Kontekst
- Metapodaci
  - Standardi metapodataka

FOSTER

# Podaci o podacima

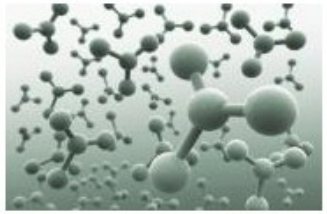
- Specifičnosti metapodatkovnih standarda
  - Znanstveno područje, instrument, kontekst ..

Search by Discipline


Biology


Earth Science


General Research Data


Physical Science


Social Science & Humanities

http://www.dcc.ac.uk/resources/metadata-standards

FOSTER

# Objava podataka - PREZERVACIJA

- Sigurnost pohranjenih podataka
- Osiguravanje dugoročne pohrane
- Planiranje mogućnosti pristupa korištenja nakon 5, 10, 100, (10000?) godina
  - Formati podataka, migracija
- Potencijalni troškovi prezervacije?
  - Problem licenciranja CC NC?

FOSTER

# Kreiranje DMP-a

- Plan upravljanja istraživačkim podacima

  *DMPOnline – Alat za kreiranje plana*

# Zaključak

- Dijeliti se mora
- Upravljanje istraživačkim podacima nikada ne smije doći u ruke komercijalnih izdavača

FOSTER

# Rasprava

FOSTER